

THE AMERICAN  
ECONOMIC REVIEW

VOL. LIV

MAY, 1964

NUMBER 3

PAPERS AND PROCEEDINGS

OF THE

*Seventy-sixth Annual Meeting*

OF THE

AMERICAN ECONOMIC ASSOCIATION

*Boston, Massachusetts, December 27-29, 1963*

*Edited by* HAROLD F. WILLIAMSON, *Secretary of the Association*  
*and*

GERTRUDE TAIT, *Executive Assistant*

PRINTED BY GEORGE BANTA COMPANY, INC.

*Publication Office:* Curtis Reed Plaza, Menasha, Wisconsin

*Executive Office:* American Economic Association  
629 Noyes Street, Evanston, Illinois

Inquiries and other communications regarding membership, meetings, and the general affairs of the Association, as well as orders for publications, should be addressed to Harold F. Williamson, Secretary of the American Economic Association, Northwestern University, 629 Noyes Street, Evanston, Illinois.

Changes of address should be sent to the Secretary at the Executive Office at Northwestern University, Evanston, Illinois. To be effective, notice must reach the office by the 1st of the month previous to the month of publication. The Association cannot be responsible for copies lost due to failure to report change of address in time.

Entered at the post office at Menasha, Wisconsin, as second class matter. Acceptance for mailing at special rate of postage under the provisions of Sec. 34-40 Par. (D) provided for in the Act of February 28, 1925, embodied in paragraph 4, section 412, P. L. and R., authorized September 13, 1928.

The *American Economic Review* is published six times a year, in January, March, May, June, September, and December, and is sent to all members of the American Economic Association as one of the privileges of membership, \$6.00 of the annual membership dues of \$8.00 being in payment of a year's subscription to the publication. Printed in U.S.A.

PRICE, \$4.00



## TABLE OF CONTENTS

	PAGE
Program of the Seventy-sixth Meeting .....	ix
PAPERS	
Richard T. Ely Lecture	
Economic Growth as an Objective of Government Policy <i>James Tobin</i>	1
Discussion	
<i>Harry G. Johnson</i> .....	21
<i>Herbert Stein</i> .....	24
The Theory of Monopolistic Competition after Thirty Years	
The Impact on Industrial Organization <i>Joe S. Bain</i> .....	28
The Impact on General Theory <i>Robert L. Bishop</i> .....	33
Monopolistic Competition and Welfare Economics <i>William J. Baumol</i>	44
Discussion	
<i>Jesse W. Markham</i> .....	53
<i>Peter O. Steiner</i> .....	55
Principles of Efficiency	
The Measurement of Waste <i>Arnold C. Harberger</i> .....	58
Efficient Allocation of Capital in an Uncertain World <i>Jack Hirshleifer</i>	77
Discussion	
<i>Dale W. Jorgenson</i> .....	86
<i>William Vickrey</i> .....	88
<i>Tjalling C. Koopmans</i> .....	92
<i>Paul A. Samuelson</i> .....	93
Efficiency in Agriculture	
Efficiency of the Farm Firm <i>Earl O. Heady</i> .....	97
Factor Market Efficiency for Agriculture <i>G. S. Tolley and B. M. Farmer</i>	107
Discussion	
<i>K. L. Robinson</i> .....	120
Efficiency in the Labor Markets	
Relative Employment Effects of Unionism <i>H. G. Lewis</i> .....	123
Labor Mobility and Wage Inflexibility <i>Robert L. Raimon</i> .....	133
The Featherbedding Problem <i>Paul A. Weinstein</i> .....	145
Discussion	
<i>William G. Bowen</i> .....	153
<i>George H. Hildebrand</i> .....	155

	PAGE
<b>The Regulated Industries</b>	
The Effect of Rate Regulation on Resource Allocation in Transportation <i>George Wilson</i> .....	160
Direct Regulation and Market Performance in the American Economy <i>Richard E. Caves</i> .....	172
The Effectiveness of Economic Regulation: A Legal View <i>Roger C. Cramton</i> .....	182
Discussion <i>Ernest W. Williams, Jr.</i> .....	192
<i>Ronald H. Coase</i> .....	194
<b>Efficiency in the Use of Natural Resources</b>	
Attainment of Efficiency in Satisfying Demands for Water Resources <i>Irving K. Fox and O. C. Herfindahl</i> .....	198
The Marine Fisheries: A Problem in International Cooperation <i>James Crutchfield</i> .....	207
Efficiency of Resource Use in Crude Petroleum: Abstract <i>M. A. Adelman</i> .....	219
Discussion <i>Alfred E. Kahn</i> .....	221
<i>Donald J. White</i> .....	223
<i>Nathaniel Wollman</i> .....	224
<b>Efficiency in the Government Sector</b>	
Fiscal Institutions and Efficiency in Collective Outlay <i>James M. Buchanan</i> .....	227
The Structure of Government and Public Investment <i>Julius Margolis</i> .....	236
Divergences between Individual and Total Costs within Government <i>Roland N. McKean</i> .....	243
Discussion <i>Mancur Olson, Jr.</i> .....	250
<i>Nathan Rosenberg</i> .....	251
<i>Jerome Rothenberg</i> .....	254
<b>Problems in Taxation</b>	
Direct Shifting of the Corporation Income Tax in Manufacturing <i>Challis A. Hall, Jr.</i> .....	258
The Tax Structure and Corporate Dividend Policy <i>John A. Brittain</i> .....	272
Economic Effects of the Capital Gains Tax <i>Martin David</i> .....	288
Discussion <i>Richard A. Musgrave</i> .....	300
<i>John Lintner</i> .....	302
<i>E. Cary Brown</i> .....	306

**Financial Factors in Business Cycles**

Financial Markets in Business Cycles: A Simulation Study <i>Frank de Leeuw</i> .....	309
Longer Waves in Financial Relations: Financial Factors in the More Severe Depressions <i>Hyman P. Minsky</i> .....	324
Discussion <i>James Duesenberry</i> .....	336
<i>John G. Gurley</i> .....	338
<i>Allan H. Meltzer</i> .....	340

**Reappraisals in American Economic History**

A New Look at Hunter's Hypothesis about the Antebellum Iron In- dustry <i>Peter Temin</i> .....	344
Antebellum Interregional Trade Reconsidered <i>Albert Fishlow</i> .....	352
Canals and Development: A Discussion of the Issues <i>Roger L. Ransom</i> .....	365
Discussion <i>Robert W. Fogel</i> .....	377

**Comparative Costs and Economic Development**

Comparative Costs and Economic Development: The Experience of India <i>Wilfred Malenbaum</i> .....	390
Comparative Costs and Economic Development: The African Case <i>Walter A. Chudson</i> .....	400
Import Substitution and Industrialization in Brazil <i>Werner Baer</i> and <i>Isaac Kerstenetzky</i> .....	411
Discussion <i>Albert O. Hirschman</i> .....	426
<i>Wolfgang F. Stolper</i> .....	428
<i>Raymond Vernon</i> .....	434

**Canadian-American Economic Relations: Efficiency and Independence**

The Canadian Tariff and the Efficiency of the Canadian Economy <i>H. C. Eastman</i> .....	437
The Exports of American-owned Enterprises in Canada <i>Albert E. Safarian</i> .....	449
The U.S. Tariff and the Efficiency of the U.S. Economy <i>Robert M. Stern</i> .....	459
Discussion <i>Lawrence B. Krause</i> .....	471
<i>C. P. Kindleberger</i> .....	474
<i>John H. Young</i> .....	477

	PAGE
<b>Efficiency of the Soviet Economy</b>	
✓The Static Efficiency of the Soviet Economy <i>Joseph S. Berliner</i> . . . .	480
✓The Dynamic Efficiency of the Soviet Economy <i>Bela Balassa</i> . . . .	490
International Division of Labor in CEMA: Limited Regret Strategy <i>Egon Neuberger</i> . . . . .	506
Discussion	
<i>Judith Thornton</i> . . . . .	516
<i>Evsey D. Domar</i> . . . . .	517
<i>Frederic L. Pryor</i> . . . . .	520
<b>Invited Dissertations, I</b>	
The Monetary Mechanism: Some Partial Relationships <i>James L. Pierce</i> . . . . .	523
The Term Structure of Interest Rates <i>Burton G. Malkiel</i> . . . . .	532
Commercial Bank Portfolio Adjustments <i>William R. Russell</i> . . . .	544
Discussion	
<i>Martin J. Bailey</i> . . . . .	554
<i>Arthur Okun</i> . . . . .	554
<i>Richard E. Quandt</i> . . . . .	557
<b>Invited Dissertations, II</b>	
The Determinants of Volume in a Probabilistic Market <i>Frederick T. Sparrow</i> . . . . .	559
The Federal Rivers and Harbors Program: An Analysis of Regional Impact <i>Robert Haveman</i> . . . . .	568
Farmer Response to Price in a Subsistence Economy: The Case of West Pakistan <i>Walter P. Falcon</i> . . . . .	580
Discussion	
<i>Charles C. Holt</i> . . . . .	592
<i>D. Gale Johnson</i> . . . . .	593
<b>Efficiency in the Teaching of Economics: The Product</b>	
The Problem of Communication <i>Leonard S. Silk</i> . . . . .	595
Economic Analysis and Public Policy: The Growing Hiatus <i>Robert R. Nathan</i> . . . . .	610
Discussion	
<i>Joseph A. Kershaw</i> . . . . .	623
<i>Kenyon A. Knopf</i> . . . . .	625
<i>Henry H. Villard</i> . . . . .	628

## PROCEEDINGS

Minutes of the Business Meeting . . . . .	635
John Bates Clark Award . . . . .	637
Minutes of the Executive Committee Meetings . . . . .	638

**Reports**

Secretary <i>Harold F. Williamson</i> .....	643
Treasurer <i>Harold F. Williamson</i> .....	649
Finance Committee <i>C. Wells Farnham</i> .....	654
Auditor <i>David Himmelblau &amp; Company</i> .....	658
Managing Editor <i>John G. Gurley</i> .....	665
Committee on Research and Publications <i>Irving B. Kravis</i> .....	670
Journal of Economic Abstracts <i>Arthur Smithies</i> .....	671
Committee on the Future Role of the A.E.A. in Economic Education <i>Ben W. Lewis</i> .....	672
Committee on the International Economic Association <i>Howard S. Ellis</i> .....	673
Advisory and Policy Board Representatives <i>Theodore Morgan</i> ....	676
Census Advisory Committee <i>Solomon Fabricant</i> .....	678
Representative to the National Bureau of Economic Research <i>Willard L. Thorp</i> .....	679
Publications of the American Economic Association .....	681

PROGRAM OF THE SEVENTY-SIXTH ANNUAL MEETING OF THE  
AMERICAN ECONOMIC ASSOCIATION

Boston, Massachusetts, December 27-29, 1963

The general theme of the sessions was the efficiency of economic systems and in particular of important parts of the American economy.

In the preparation of the program, the President received substantial assistance from the following members of the Association: Gregory Grossman, O. C. Herfindahl, Bert Hickman, D. Gale Johnson, Harry G. Johnson, Dale Jorgenson, C. P. Kindleberger, H. G. Lewis, Roland N. McKean, Gerald M. Meier, Douglass North, and Joseph Pechman.

Thursday, December 26, 1963

6:00 P.M.

*Executive Committee Dinner Meeting*

Friday, December 27, 1963

9:30 A.M.

*Principles of Efficiency*

*Chairman:* PAUL A. SAMUELSON, Massachusetts Institute of Technology

*Papers:* ARNOLD HARBERGER, University of Chicago; JACK HIRSCHLEIFER, University of California, Los Angeles

*Discussants:* DALE JORGENSEN, University of California, Berkeley; WILLIAM VICKREY, Columbia University; TJALLING C. KOOPMANS, Yale University; PAUL A. SAMUELSON, Massachusetts Institute of Technology

*Reappraisals in American Economic History*

*Chairman:* DOUGLASS NORTH, University of Washington

*Papers:* PETER TEMIN, Harvard University; ALBERT FISHELOW, University of California, Berkeley; ROGER RANSOM, University of Virginia

*Discussant:* ROBERT FOGEL, University of Chicago

*Efficiency in Agriculture* (Joint session with the American Farm Economic Association)

*Chairman:* D. GALE JOHNSON, University of Chicago

*Papers:* EARL HEADY, Iowa State University; GEORGE TOLLEY and B. H. FARMER, North Carolina State College

*Discussants:* GLENN L. JOHNSON,<sup>1</sup> Michigan State University; WILLARD COCHRANE,<sup>1</sup> U.S. Department of Agriculture; KENNETH L. ROBINSON, Cornell University

2:30 P.M.

*Comparative Costs and Economic Development*

*Chairman:* C. P. KINDLEBERGER, Massachusetts Institute of Technology

*Papers:* WILFRED MALENBAUM, University of Pennsylvania; WALTER A. CHUDSON, United Nations; WERNER BAER and ISAAC KERSTENETZKY, Yale University

*Discussants:* ALBERT HIRSCHMAN, Columbia University; WOLFGANG STOLPER, University of Michigan; RAYMOND VERNON, Harvard University

*Problems in Taxation*

*Chairman:* JOSEPH PECHMAN, Brookings Institution

*Papers:* CHALLIS HALL, Yale University; JOHN BRITTAIN, Vanderbilt University; MARTIN DAVID, University of Wisconsin

*Discussants:* RICHARD MUSGRAVE, Princeton University; JOHN LINTNER, Harvard University; E. CARY BROWN, Massachusetts Institute of Technology

*Efficiency in the Labor Markets* (Joint session with Industrial Relations Research Association)

*Chairman:* HAROLD M. LEVINSON, University of Michigan

*Papers:* H. G. LEWIS, University of Chicago; R. L. RALMON, Cornell University; P. A. WEINSTEIN, Columbia University

*Discussants:* WILLIAM G. BOWEN, Princeton University; GEORGE HILDEBRAND, Cornell University; CHARLES KILLINGSWORTH, Michigan State University

<sup>1</sup>No manuscript received.

8:00 P.M.

*Presidential Address<sup>2</sup>*

*Chairman:* EDWARD S. MASON, Harvard University  
*Address:* GOTTFRIED HABERLER, Harvard University

Saturday, December 28, 1963

9:30 A.M.

*Efficiency in the Government Sector*

*Chairman:* C. LOWELL HARRISS, Columbia University  
*Papers:* JAMES M. BUCHANAN, University of Virginia; JULIUS MARGOLIS, University of California, Berkeley; ROLAND N. MCKEAN, RAND Corporation  
*Discussants:* MANCUR OLSON, JR., Princeton University; NATHAN ROSENBERG, Purdue University; JEROME ROTHENBERG, Northwestern University

*The Regulated Industries*

*Chairman:* MERRILL ROBERTS, University of Pittsburgh  
*Papers:* GEORGE WILSON, Indiana University; RICHARD CAVES, Harvard University; ROGER CRAMTON, University of Michigan  
*Discussants:* ERNEST WILLIAMS, Columbia University; RONALD COASE, University of Virginia

*Efficiency of the Soviet Economy*

*Chairman:* GREGORY GROSSMAN, University of California, Berkeley  
*Papers:* JOSEPH S. BERLINER, Brandeis University; BELA BALASSA, Yale University; EGON NEUBERGER, RAND Corporation  
*Discussants:* JUDITH THORNTON, University of Washington; EVSEY D. DOMAR, Massachusetts Institute of Technology; FREDERIC L. PRYOR, University of Michigan

*The Poverty Problem in American Agriculture (Joint session with the American Farm Economic Association)<sup>3</sup>*

*Chairman:* STANLEY K. SEAVER, University of Connecticut  
*Papers:* LEE MARTIN, New England Board of Higher Education; LOUIS UPCHURCH, U.S. Department of Agriculture  
*Discussants:* EDWARD J. BOOTH, Oklahoma State University; J. W. WILEY, Purdue University

12:30 P.M.

*Joint Luncheon with the American Finance Association<sup>4</sup>*

*Chairman:* GEORGE T. CONKLIN, American Finance Association  
*Speaker:* ROBERT V. ROOSA, Undersecretary of the Treasury

2:30 P.M.

*Efficiency in the Teaching of Economics: The Product*

*Chairman:* BEN W. LEWIS, Oberlin College  
*Papers:* LEONARD S. SILX, Business Week; ROBERT R. NATHAN, Robert R. Nathan Associates  
*Discussants:* JOSEPH A. KERSHAW, Williams College; KENYON A. KNOPF, Grinnell College; HENRY H. VILLARD, New York City College

*Canadian-American Economic Relations: Efficiency and Independence*

*Chairman:* V. W. Bladen, University of Toronto  
*Papers:* H. C. EASTMAN, University of Toronto; A. E. SAFARIAN, University of Saskatchewan; ROBERT M. STERN, University of Michigan  
*Discussants:* LAWRENCE KRAUSE, Brookings Institution; C. P. KINDLEBERGER, Massachusetts Institute of Technology; J. H. YOUNG, University of British Columbia

*The Theory of Monopolistic Competition After Thirty Years*

*Chairman:* JOSEPH J. SPENGLER, Duke University  
*Papers:* JOE BAIN, University of California, Berkeley; ROBERT BISHOP, Massachusetts Institute of Technology; WILLIAM BAUMOL, Princeton University  
*Discussants:* P. O. STEINER, University of Wisconsin; JESSE MARKHAM, Princeton University

<sup>2</sup> Published in the March, 1963, *A.E.R.*

<sup>3</sup> Published in the May, 1963, *J. of Farm Econ.*

<sup>4</sup> Published in the March, 1963, *J. of Fin.*

8:00 P.M.

*Richard T. Ely Lecture*

*Chairman:* SIMON KUZNETS, Harvard University

*Paper:* JAMES TOBIN, Yale University

*Discussants:* HARRY G. JOHNSON, University of Chicago; HERBERT STEIN, Committee for Economic Development

Sunday, December 29, 1963

9:30 A.M.

*Invited Dissertations, I*

*Chairman:* WILLIAM FELLNER, Yale University

*Dissertations:* J. L. PIERCE, Yale University; BURTON MALKIEL, Princeton University; WILLIAM RUSSELL, University of Wisconsin

*Discussants:* MARTIN BAILEY, University of Chicago; ARTHUR OKUN, Yale University; RICHARD QUANDT, Princeton University

*Efficiency in the Use of Natural Resources*

*Chairman:* EDWARD S. MASON, Harvard University

*Papers:* IRVING K. FOX and O. C. HERFINDAHL, Resources for the Future, Inc.; JAMES CRUTCHFIELD, University of Washington; M. A. ADELMAN, Massachusetts Institute of Technology

*Discussants:* ALFRED E. KAHN, Cornell University; DONALD WHITE, Boston College; NATHANIEL WOLLMAN, University of New Mexico

*Local and Regional Impacts of Shifts in Military Expenditures* (Joint session with the Regional Science Association)

*Chairman:* GEORGE ELLIS, Federal Reserve Bank of Boston

*Papers:* CHARLES TIEBOUT, Washington, and RICHARD PETERSON, Bank of America; WALTER ISARD, University of Pennsylvania, and EUGENE W. SCHOOLER, West Chester State College

*Discussants:* GEORGE A. STEINER, University of California, Los Angeles; WILBUR THOMPSON, Wayne State University

*Intertemporal Economic Theory* (Joint session with the Econometric Society)

*Chairman:* WASSILY W. LEONTIEF, Harvard University

*Papers:* ROBERT DORFMAN, Harvard University; TJALLING KOOPMANS, Yale University; ROY RADNER, University of California, Berkeley

*Discussants:* SUKHAMOY CHAKRAVARTY, New Delhi School of Economics; PETER A. DIAMOND, University of California, Berkeley; EDMUND S. PHELPS, Yale University

2:30 P.M.

*Invited Dissertations, II*

*Chairman:* FRANCO MODIGLIANI, Massachusetts Institute of Technology

*Dissertations:* F. T. SPARROW, Johns Hopkins University; ROBERT HAVEMAN, Grinnell College; WALTER FALCON, Harvard University

*Discussants:* CHARLES HOLT, University of Wisconsin; ROBERT DORFMAN,<sup>1</sup> Harvard University; D. GALE JOHNSON, University of Chicago

*Financial Factors in Business Cycles*

*Chairman:* BERT HICKMAN, Brookings Institution

*Papers:* FRANK DE LEEUW, Board of Governors of the Federal Reserve System; HYMAN P. MINSKY, University of California, Berkeley

*Discussants:* JAMES DUESENBERY, Harvard University; JOHN G. GURLEY, Stanford University; ALLAN H. MELTZER, Carnegie Institute of Technology

5:00 P.M.

*Business Meeting*

6:00 P.M.

*Executive Committee Dinner Meeting*



THE purpose of the American Economic Association, according to its charter, is the encouragement of economic research, the issue of publications on economic subjects, and the encouragement of perfect freedom of economic discussion. The Association as such takes no partisan attitude, nor does it commit its members to any position on practical economic questions. It is the organ of no party, sect, or institution. Persons of all shades of economic opinion are found among its members, and widely different issues are given a hearing in its annual meetings and through its publications. The Association, therefore, assumes no responsibility for the opinions expressed by those who participate in its meetings. Needless to say, the papers presented are the personal opinions of the authors and do not commit the organizations or institutions with which they are associated.

HAROLD F. WILLIAMSON  
*Secretary*

RICHARD T. ELY LECTURE

ECONOMIC GROWTH AS AN OBJECTIVE  
OF GOVERNMENT POLICY\*

By JAMES TOBIN  
*Yale University*

In recent years economic growth has come to occupy an exalted position in the hierarchy of goals of government policy, both in the United States and abroad, both in advanced and in less developed countries, both in centrally controlled and decentralized economies. National governments proclaim target growth rates for such diverse economies as the Soviet Union, Yugoslavia, India, Sweden, France, Japan—and even for the United Kingdom and the United States, where the targets indicate dissatisfaction with past performance. Growth is an international goal, too. The Organization for Economic Cooperation and Development aims at a 50 percent increase in the collective gross output of the Atlantic Community over the current decade.

Growth has become a good word. And the better a word becomes, the more it is invoked to bless a variety of causes and the more it loses specific meaning. At least in professional economic discussion, we need to give a definite and distinctive meaning to growth as a policy objective. Let it be neither a new synonym for good things in general nor a fashionable way to describe other economic objectives. Let growth be something it is possible to oppose as well as to favor, depending on judgments of social priorities and opportunities.

I

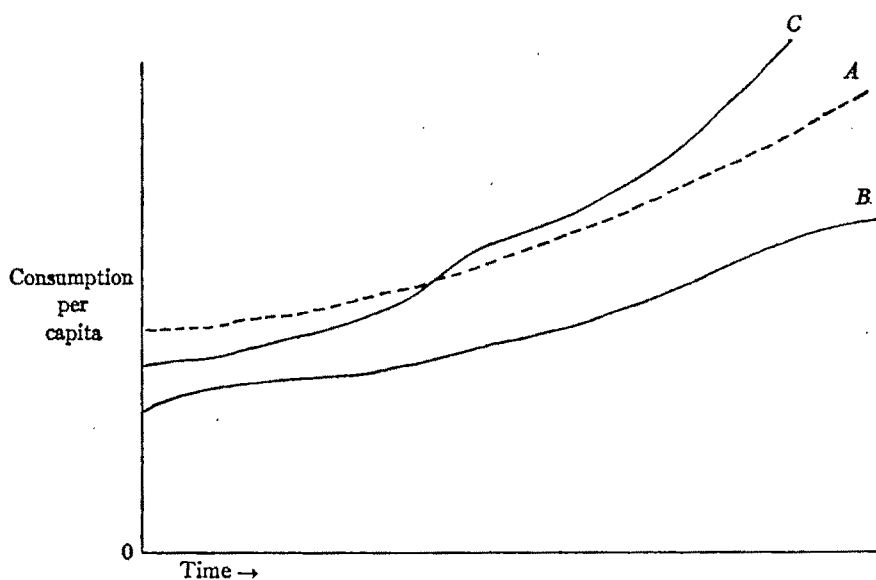
In essence the question of growth is nothing new, but a new disguise for an age-old issue, one which has always intrigued and preoccupied economists: the present versus the future. How should society divide its resources between current needs and pleasures and those of next year, next decade, next generation?

The choice can be formalized in a way that makes clear what is essentially at stake. A consumption path or program for an economy describes its rate of consumption at every time point beginning now and extending indefinitely into the future. Not all imaginable consumption paths are feasible. At any moment future possibilities are limited by

\* I am greatly indebted to my colleagues at the Cowles Foundation, especially Tjalling Koopmans, Arthur Okun, and E. S. Phelps, for clarifying many of the questions discussed in this paper. But they do not necessarily share my opinions, and they certainly share no responsibility for my mistakes.

our inherited stocks of productive resources and technological knowledge and by our prospects for autonomous future increase in these stocks. Of feasible paths, some dominate others; i.e., path A dominates B if consumption along path A exceeds consumption along path B at every point of time. I hope I will incur no one's wrath by asserting that in almost everyone's value scheme more is better than less (or certainly not worse), at least if we are careful to specify more or less of

#### ALTERNATIVE CONSUMPTION PATHS



A dominates B but not C.

FIGURE 1

what. If this assertion is accepted, the interesting choices are between undominated or efficient feasible paths; e.g., between a pair A and C where A promises more consumption at some points in time but less at others. See Figure 1. In particular, I take growthmanship to be advocacy of paths that promise more consumption later in return for less earlier.

But growthmanship means more than that. Growthmen are usually willing to throw the weight of the government on to the scales in order to tip the balance in favor of the future. Here they fly in the face of a doctrinal tradition of considerable strength both in economics and in popular ideology. Does not the market so coordinate the free, decentralized decisions of individuals between present and future so as to reach an optimal social choice? Is not any government intervention in favor of growth, therefore, bound to tilt the scales toward the future to a degree that society does not "really" want?

The basic question raised by advocates of faster growth may be further formalized to emphasize this issue. Assuming that the economy is now on a feasible and undominated consumption path, the desirability of deviating from it can be expressed in the language of interest rates and present values. Any feasible and efficient path, including the prevailing path, implies two sets of interest rates. One, which we may call the time preference set, expresses the society's marginal rates of substitution as consumers between consumption at one date and consumption at another date. This set answers questions like the following: Given society's consumption prospect, how much increase in consumption five years or fifty years or  $t$  years from now is worth the loss of a dollar's worth of consumption today? The rates implied by the answers need not all be the same. The other set, which we may call the technological set, expresses the opportunities which present and prospective technology offers the society for marginal substitutions of consumption at one date for consumption at another. This set answers questions like the following: Given the consumption path, by how much could consumption be increased five years or fifty years or  $t$  years from now by the resources released from a dollar's worth of consumption today? Again, the rates can vary with time. A sacrifice in current consumption may yield, say, 10 percent per year if its fruits are taken five years from now, but 20 percent—or 2 percent—if they are taken fifty years from now.

A small proposed feasible deviation from the existing path can in principle be tested as follows: Calculate the present values of the proposed deviations in consumption, negative and positive, discounting them by the time preference set of interest rates. If the sum is positive, the proposed deviation is worth while. If it is zero or negative, it is not worth while. We know that this sum will not be positive if it happens that the time preference and technological interest rates are identical.

Evidently growthmen believe that the two sets diverge in such a way that society would give a positive present value to feasible increases in future consumption purchased at the expense of current and near future consumption. Their opponents think the contrary. Many of them have faith in the capital markets and believe there is a presumption that these markets make the two sets of rates equal.

## II

This is the heart of the issue, I believe, and I shall return to it later in this lecture. First, however, I must discuss some questions raised by the formulation of the growth issue which I have just tried to sketch. What is the relationship between growth and other objectives of economic policy, in particular full employment of resources? Are there some noneconomic reasons for accelerating growth—reasons which this

formulation excludes or evades? Exactly what is the "consumption" whose path is to be chosen? Finally, can government successfully influence the growth path?

1. *Growth Versus Full Employment.* To accelerate growth is not the same thing as to increase the utilization of existing resources, manpower, and capital capacity. In the formulation sketched above, a consumption path with underutilization is dominated or inefficient. By putting the idle resources to work, consumption can be increased both now and in future. The same is true of other measures to improve the efficiency of allocation of resources. We can all agree, I presume, on the desirability of growth measures free of any cost. If that is the meaning of growth policy, there is no issue.

For short periods of time, stepping up the utilization of capacity can increase the recorded rate of growth of output and consumption. But over the decades fluctuations in the utilization of capacity will have a minor influence compared to the growth of capacity itself. To express the same point somewhat differently, the subject of economic growth refers mainly to supply, or capacity to produce, rather than to demand. In the short run, accelerating the growth of demand for goods and services can, by increasing the rate of utilization of capacity, speed the growth of output. But in the long run, output and real demand cannot grow faster than capacity. If monetary demand is made to set a faster pace, it will be frustrated by a rate of inflation that cuts real demand down to size.

Public policy affecting aggregate demand should be aimed at maintaining a desired rate of utilization of capacity. Economists and other citizens will differ on how high this rate should be, because they differ in the weights they attach to additional employment and output, on the one hand, and to the risks of faster price inflation, on the other. But however this balance is struck, monetary and fiscal policies can in principle hit the target utilization rate just as well whether the economy's capacity is growing at 5 percent or 3 percent or zero percent.

Full employment is, therefore, not a reason for faster economic growth; each is an objective in its own right. In an economy suffering from low rates of utilization of manpower and capital resources, accelerating the growth of aggregate demand may well be the need of the hour. But this ought not be considered growth policy in the more fundamental sense. Tax reduction today has sufficient justification as a means of expanding demand and raising the rate of utilization. It is probably an unfortunate confusion to bill it as a growth measure, too.

I do not mean, of course, that the rate of growth of the economy's capacity is in practice wholly independent of its rate of utilization. In principle they may be independent. Demand can be expanded in ways

that do not accelerate, indeed may even retard, the growth in capacity itself. But as a rule some of the output resulting from an increase in utilization will be used in ways that expand future capacity. Thus the Great Depression deprived the nation and the world of investment as well as consumption; we, as well as our fathers, bear the cost. The proposed tax reduction, even though its major impact is to stimulate consumption, will nonetheless increase the share of national capacity devoted to capital accumulation. It is in this sense that it can be called a growth measure. But there may be ways to expand demand and utilization to the same degree while at the same time providing both more stimulus for and more economic room for capacity-building uses of resources now idle.

2. *Noneconomic Reasons for Growth.* Economic growth may be a national objective for noneconomic reasons, for national prestige or national strength or national purpose.

No doubt much of the recent dissatisfaction with U.S. growth is motivated by unfavorable comparisons with other countries, especially the Soviet Union. If current rates are mechanically extrapolated, it is easy to calculate that the U.S. will not be first in international statistical comparisons in our great grandchildren's textbooks. Presumably the American nation could somehow stand and even rationalize this blow to our national pride, even as we survive quadrennial defeats by Russian hordes in the Olympics. At any rate, it is not for professional economists to advise the country to act differently just to win a race in statistical yearbooks. The cold war will not be so easily won, or lost, or ended.

International competition in growth may, however, be of importance in the battle for prestige and allegiance among the "uncommitted" and less developed countries. These nations place a high premium on rapid economic progress. They will not—so the argument runs—choose the democratic way in preference to communism, or market economies in preference to centrally directed economies, unless our institutions show by example that they can outperform rival systems. A political psychologist rather than an economist should evaluate this claim. But it has several apparent weaknesses: (a) Rate of growth is not the only dimension of economic performance by which our society will be judged by outside observers. Equality of opportunity and of condition, humanity, understanding, and generosity in relation to less privileged people in our own society and abroad—these are perhaps more important dimensions. (b) The U.S. is not the only noncommunist economy. The examples of Western Europe (in particular the contrast of Western to Eastern Germany) and Japan are more relevant to the rest of the world, and they give convincing evidence of the economic vitality of free societies. (c) What is much more important is a demonstration that

an underdeveloped country can progress rapidly under democratic auspices. Without this kind of demonstration, faster growth of affluence in already affluent societies may cause more disaffection than admiration.

On the score of national strength, there is a case for growth. But it is more subtle than the facile association of military power with generalized civilian economic capacity. Nuclear technology has made this connection looser than ever. A country is not necessarily stronger than another just because it has a higher GNP. Great productive capacity may have been the decisive reserve of military strength in the last two World Wars, but nowadays it is useless if it remains unmobilized until the cataclysmic buttons are pushed. A country with smaller GNP can be as strong or even stronger if it persistently allocates enough of its GNP to military purposes. And in the age of overkill, apparently there can be a point of saturation.

Should we grow faster to be better prepared to meet possible future needs for output for military purposes—or for other uses connected with national foreign policy? If we do not, we will have to meet such needs when they arise by depriving other claimants on national production, principally consumption, at the time. But in order to grow faster, we have to deprive these claimants now. Hence the national power argument seems to boil down to the economist's calculation after all; i.e., to the terms of trade between current and future consumption.

But there is an important exception. Some hazards are great enough to bias our choice to favor the future over the present, to accept less favorable payoffs than we otherwise would. We might conceivably be challenged one day to a duel of overriding priority, involving all-out commitment of resources to military uses, foreign aid, space adventures, or all of these together. A high GNP might be the difference between victory and defeat rather than the difference between more or less consumption. In other words, this contingency is one that could be met only by sacrifices of consumption in advance, not by sacrifices at the time.

As for national purpose, it is surely conceivable that a growth target could inspire, galvanize, and unite the nation. But it is not the only objective that could serve this purpose, nor is it necessarily the best candidate.

3. *Growth in What?* The formulation of the growth issue sketched above presents it as a choice among available consumption paths. The concentration on consumption deserves some elaboration and explanation—especially because growth performance and aspiration are popularly expressed in terms of gross or net national product.

Some of the noneconomic reasons for favoring faster growth also sug-

gest that GNP is the relevant measure, especially if it is the most usual and visible measure. But as economists we would make welfare or utility depend on consumption. We would require the investment part of GNP to derive its value from the future consumption it supports. After all, a future in which the rate of growth of GNP reaches fantastic heights has no appeal if the fruits of the achievement are never consumed. We must heed the "golden rule" of capital accumulation: there is a saving ratio and a corresponding capital intensity that maximize consumption. Persistent saving in excess of the rule makes GNP higher but consumption lower. (See Phelps [5].)

Neither GNP nor consumption, as ordinarily measured, counts leisure. Yet I do not understand advocates of faster growth to be taking a stand in favor of goods and services priced in the market and against leisure. Should the trend toward shorter hours, longer vacations, and earlier retirements accelerate, the rate of growth of consumption as measured in the national accounts might decline. But a decline for this reason should not bother a growth-oriented economist. The *Affluent Society* to the contrary notwithstanding, the conventional wisdom of economics was long since liberated from the fallacy that only produced goods and services yield utility and welfare. Economists do have prejudices against biasing the price system in favor of leisure and against forcing the leisure of involuntary unemployment on anyone. But those are other matters. The consumption whose growth path concerns us should include leisure valued at the real wage. Needless to say, it should also allow for consumption goods and services provided by government.

Finally, is the relevant measure aggregate consumption or consumption per capita? Later in the lecture I shall be concerned with social indifference curves between consumption at one date and at a later date. An example is pictured in Figure 2. What measure of consumption should the axes of such a diagram represent? The answer depends on questions like the following: Do we discharge our obligation to the next generation if we enable them to enjoy the same aggregate consumption even though there will be more of them to share it? Should we, on the other hand, sacrifice today in order to raise per capita consumption half a century from now just because there will then be more consumers? Or should generations count in some sense equally regardless of size?

These are not easy questions for the social philosopher, but revealed social preferences lean towards per capita consumption. Presumably we do not value increase in population for its own sake. We might if sheer numbers were important for national power. But in general we are content to leave population trends to free choice; indeed, we seek to enlarge parents' ability to limit births at their discretion. Neither immigration



nor subsidies for childbearing are advanced as growth proposals. In the world at large, certainly, the commonly accepted aim is to retard the growth of population, not to accelerate it. (For discussion of some aspects of this problem, see Koopmans [3].)

4. *Government's Power to Influence Growth.* I come now to the question whether the government can influence growth, even if we wish it to. The growth objective is commonly framed in terms of an exponential growth rate. Those who advocate measures to promote growth frequently are expressing a preference for a higher per annum rate of growth, for 4 percent or 5 percent instead of 3 percent or  $3\frac{1}{2}$  percent. But the thrust of much recent theorizing and model building is that in the really long run we have no choice about the growth rate. (See, for example, Phelps [6].) The long-run growth rates of GNP and aggregate consumption are exogenously determined by the growth of the labor force and the progress of technology. Or, to express the same conclusion somewhat differently, the rates of growth of productivity per man and

SUGGESTED CRITERION OF INTERTEMPORAL IMPARTIALITY

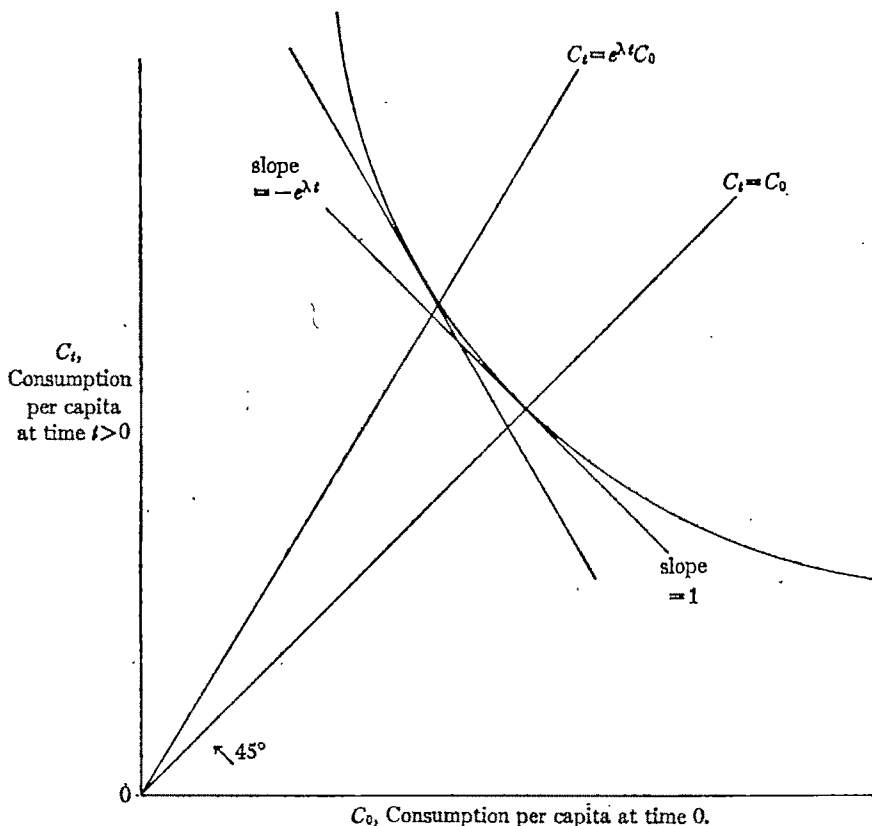


FIGURE 2

of consumption per capita are in the long run controlled by the rate of advance of technology.

According to these models, there are various hypothetical paths which share the exogenously determined rate. These paths differ in level. On a higher path, consumption per capita is always larger than on a lower one. A higher path represents a higher capital intensity (so long as capital intensity does not exceed its golden rule value), and a correspondingly higher propensity to save is required to maintain it.

An economy moving along one of these paths may "decide" to move to a higher one, by lowering its propensity to consume. For a while, its growth rate will be higher, as the effects of increasing capital intensity and modernization are added to those of the underlying progress of technology. Eventually, however, capital intensity will cease to increase and the growth rate will converge to its natural value. The process can be repeated by further increases in the saving ratio, but the golden rule argument cited above sets a limit long before the propensity to consume reaches zero—indeed, when the propensity to save is equal to the elasticity of output with respect to capital accumulation. This is the highest path for consumption per capita.

Asymptotically, then, it appears that we have no choice about our rate of growth, but can choose only between parallel paths of different levels. But asymptotically is a very long time. The period of transition from one path to another, short from the perspective of the model builder, may be measured in decades or generations. It is therefore not wholly misleading to regard society as choosing among growth rates.

Models of this kind take the rate of technological progress as exogenous. In fact, it is probably subject to improvement, like the degree of capital intensity, by expenditure of current resources. We still know very little about the technology that governs the production of applicable technological knowledge. What is required to keep the index of technology, which determines the productivity of labor and capital, growing at a constant exponential rate? Does it take simply a constant absolute amount of labor and capital? Does it take a constant fraction of the resources devoted to production? Does it take an input of resources growing at the same rate as the technology index itself? Only when we can answer such questions can we know whether and how the pace of economic growth is ultimately limited by the natural increase of the labor force.

A second reason for doubting that government measures can affect the intertemporal choices of society is the possibility that the private decisions of individuals can and will offset these measures. Suppose, for example, that the government levies new taxes and uses the proceeds for saving and investment, either through public expenditure or through public lending to private investors or through retirement of public debt.

The government's purpose is to increase later consumption at the expense of earlier. But if this purpose is perfectly well understood, will not the public reduce its private saving in the knowledge that its collective saving is now doing part of the job?

I have two comments regarding this possibility. First, it may be that the government's saving corrects a situation of underinvestment, where public or private projects that would pay for themselves in social benefits (discounted at the time preference set of interest rates) were not being undertaken. In this case, the government's twist of the path will not be undone even if perfectly understood because the new path corresponds better to public preferences. Second, the assumption that the public correctly foresees all the consequences of government policy is farfetched. In the example above, economists would usually expect the new taxes to be paid in large part out of private consumption. Disposable income is reduced; and so, gradually, is the public's net financial claim on the government—a more tangible element in private balance sheets than the present value of future tax liabilities or of free services from government.

I conclude, therefore, that at least for the medium run, government can affect the growth of the economy; and I turn to the question whether it should.

### III

In this section I propose to argue: (1) that government might legitimately have a growth policy, and indeed could scarcely avoid having one, even if private capital markets were perfect; (2) that capital markets are far from perfect and that private saving decisions are therefore based on an overconservative estimate of the social return to saving; and (3) that the terms on which even so advanced an economy as our own can trade present for future consumption seem to be very attractive.

1. *Government Neutrality in Intertemporal Choice.* Many economists and many other citizens will argue that the government should be neutral as between present and future. In their view the capital markets produce an optimal result, balancing the time preferences of individuals, freely expressed through their consumption and saving behavior, against the technological opportunities for substituting consumption tomorrow for consumption today. Let us assume for the moment that government can be neutral in some meaningful sense and that the capital markets perform their assigned function. Even so, I believe government should have a growth policy, and only by accident a neutral one.

I fail to see why economists should advise the public that it is wrong for them collectively to supplement (or diminish) the provisions for the future they are making individually. I agree to the desirability of satis-

ifying human preferences—that is what our kind of society and economy is all about. But I have never been able to understand why the preferences of individuals are worthy of respect only when they are expressed in the market, why the preferences of the very same individuals expressed politically should be regarded as distortions. Sometimes economists come close to rationalizing all market results and private institutions by the argument that they would not occur and survive if they were not optimally satisfying individuals' preferences. But political results and public institutions are not granted the benefit of presumptive justification-through-existence.

In both arenas preferences certainly need to be guided by full and accurate information. In the arena of government policy, it is the business of economists to help the society know what it is doing, to understand the choices, benefits, costs, and risks it confronts, not simply to repeat *ad nauseum* that the best thing to do is nothing.

The case for explicit government policy in intertemporal social choice is especially strong. More than any other social institution, government represents the permanence and continuity of the society. And in democracy one way in which each generation uses government is to protect the interests of unborn generations against its own shortsightedness and selfish instincts.

We cannot be sure that lineal family ties will give individual sufficient motivation to provide for society's future. Suppose the individuals of a whole generation, deciding that their children and grandchildren might better start from scratch, were to proceed to consume their capital. Good capital markets might reflect this epidemic of acute time preference in a perfectly Pareto-optimal way. But would we as a nation feel that we were collectively discharging our obligations to our successors?

Through many activities of government, including conservation and public education, we have recognized a generalized obligation to equip the next generation—an obligation wholly distinct from our individual provisions for our own children. This generalized obligation acquires special force if we take seriously our ideals of equality of opportunity. We like to think that our society gives the members of each generation an equal chance in the race, or at least that their chances are not predetermined by family backgrounds. Besides requiring investment in human beings on a basis other than ability to pay, this ideal suggests redistributive taxation of estates. And if estate taxation dulls incentive to save for specific heirs, the government needs to replenish saving collectively.

But what is growth-neutral government finance anyway? I have already dismissed as farfetched one answer; namely, that any government finance is growth neutral when it is fully and accurately foreseen

and accordingly offset, by taxpayers and by the beneficiaries of government services. Often a balanced budget is considered a growth-neutral fiscal policy. The budget in this rule is not, of course, the conventional U.S. administrative budget. Rather the rule suggests that (a) net government investment should be covered by borrowing, with the Treasury competing in the capital markets with private investors for private saving, and that (b) other government expenditure, including allowance for consumption of public capital, should be covered by current taxes or fees.

The rule is clear cut and has intuitive appeal. But it seems to bias social choice against the future when there is simply a shift in public preference from private consumption, present and future, to collective consumption, present and future. The rule would levy only enough new taxes to cover the additional collective consumption. But the evidence is that taxpayers would pay some of these new taxes from saving (especially if the collective consumption the taxes finance were of regrettable necessities like national defense rather than of services that clearly yield utility now and in future). Interest rates would rise and investment would be curtailed, even though no shift in social time preference has occurred. Clearly the 10 percent of GNP which the cold war has forced us to devote persistently to national defense has not come wholly from private or public consumption. True neutrality evidently would require a tighter fiscal policy the bigger the government's budget for current consumption.

But in any case, the quest for neutrality is probably a search for a will-of-the-wisp. For it is not only the overall budget position of government but also the specifics of taxation and expenditure which affect intertemporal choices. We have not yet learned how to implement the welfare economist's lump-sum taxes. I have already given one example of a tax which is desirable in view of other social objectives but is bound to affect incentives for private accumulation of wealth. It will suffice to remind you also that our methods of taxation necessarily favor one kind of current consumption, leisure, both as against other current consumption and as against future consumption of products and leisure.

The major policy proposals of growthmen boil down to the suggestion that government should save—or save more—by making investments on its own account, subsidizing the investments of others, or by channeling tax money through the capital markets into private investment. This last item is the major purpose of the full employment budget surplus for which Councils of Economic Advisers longed under both Presidents Eisenhower and Kennedy.

It is now widely recognized that in principle the government can match aggregate demand to the economy's capacity in a variety of ways. Its various instruments for regulating or stabilizing demand affect con-

sumption and investment differently. A strong pro-growth policy would restrict consumption by taxation or by economy in government's current expenditure while stepping up public investment and encouraging private investment through tax incentives or low interest rates and high liquidity. The government cannot avoid choosing some combination of its demand-regulating instruments. Therefore government is bound to affect the composition of current output and society's provision for the future. Let us debate this choice of policy mixtures on its merits, weighing growth against its costs and against other objectives of policy, without encumbering the debate with a search for that combination which meets some elusive criterion of neutrality.

*2. Imperfections in Private Capital Markets.* I turn now to the second subject: the efficiency of the capital markets. Do private saving decisions reflect the real payoffs which nature and technology offer the economy? There are several reasons to believe that the answer is negative.

*Monopoly and Restrictions of Entry.* The evidence is that the rates of return required of real investment projects by U.S. business corporations are very high—typically more than 10 percent after allowance for depreciation, obsolescence, and taxes. Rates of this magnitude are not only required *ex ante* but realized *ex post*. Why do these rates so greatly exceed the cost of borrowed funds, the earnings-to-price ratio of equity issues, and in general the rates of return available to savers?

One reason clearly is that the relevant markets are not purely competitive. A monopolistic or oligopolistic firm limits its expansion in product markets, its purchases in factor markets, and its calls on capital markets, because the firm takes into account that prices and rates in these markets will turn against it. The managers seek to maintain a market valuation of the firm in excess of the replacement cost of its assets, the difference representing the capitalized value of its monopoly power, often euphemistically called good will. Restrictions and costs of entry prevent other firms from competing this difference away. Foresighted and lucky investors receive the increases in the firm's market value in the form of capital gains. But the willingness of savers to value the assets of the firm above their cost, i.e., to supply capital at a lower rate of return than the firm earns internally, is not translated into investment either by this firm or by others. One effect is to depress rates of return in more competitive sectors of the economy. But another result is to restrict total saving and investment.

*Risks, Private and Social.* Risks provide a second reason for the observed divergence between the rates of return satisfactory to savers and those typically required of real investment projects. Some of these are risks to the economy as well as to the owners of the business: technological hazards, uncertainties about consumer acceptance of new products, or uncertainties about the future availability and social opportunity

cost of needed factors of production. Even though these are social as well as private risks, it is not clear that society should take a risk-averse position towards them and charge a risk premium against those projects entailing more uncertainties than others. Presumably society can pool such risks and realize with a very small margin of uncertainty the actuarial return on investments.

Moreover, some of the private risks are not social risks at all. Consider, for example, uncertainties about competition and market shares; if several rivals are introducing a new process or new product, the main uncertainties in the investment calculation of each are the future actions of the others. Consider, further, the high and sometimes prohibitive cost which many firms impute to external funds—apparently as insurance against loss of control to new shareowners, or, with extremely bad luck, to bondholders. If savers were offered the rates of return asked of and earned by business investments, in the form of assets that impose no more risk on the holder than is commensurate to the social risks involved, presumably they would choose to save more.

It is true, on the other hand, that some net saving is now motivated by personal contingencies that are likewise social risks of a much smaller order. But our society has created insuring institutions, both private and public, to reduce the need for oversaving to meet such contingencies. Except in the field of residential construction, it has created few similar institutions to prevent private risk-aversion from leading to under-investment.

*External Returns to Investment.* Some investments yield benefits which cannot be captured by the individual or firm making the initial outlay. Research and development expenditures and outlays for training of personnel are obvious cases in point. Government policy has already recognized this fact both in tax law and in government expenditures, and it is difficult to judge whether this recognition is sufficient. Kenneth Arrow [1] has pointed out that not only R and D but all forms of investment activity share in some degree the property that B may learn from A's doing. The support which this observation gives to a general policy of encouraging investment is somewhat tempered by reflecting that the same social process of "learning by doing" can occur in production of goods and services for current consumption. However, experience is most important as a teacher in new situations, and innovations are likely to require investment.

In regard to investment in human capacities and talents, it is by no means clear that public outlays are yet sufficient to reap the external benefits involved, or even that the relevant capital markets are sufficiently developed to permit individuals to earn the private benefits. I recognize that calculations of the rate of return to educational outlays depend critically on how much of these outlays are charged to current

consumption. As an educator and ex-student I am inclined to rate high the immediate utility-producing powers of education.

3. *The Payoff to Social Saving.* The burden of my remarks so far is that we cannot escape considering growth or, more precisely, intertemporal choice as an issue of public economic policy. We cannot assume, either, that the market settles the issue optimally or that government can be guided by some simple rules of neutrality. We—and here I mean the economics profession and the country and not the three of us speaking tonight—must confront head-on the question whether the social payoff of faster growth in higher future consumption validates its cost in consumption foregone today. The issue that needs to be joined is typified by the contrast between Denison [2], who estimates a very high investment requirement for a one point increase in the medium-term growth rate (a ten point increase in the ratio of current gross investment to GNP) and Solow [8], who calculates a marginal investment requirement only about one-fifth as high.

Fortunately the profession has now begun the task of computing rates of return on various kinds of investment, tangible and intangible. Thanks to theoretical advances in growth models and in handling the knotty problems of technological progress, vintage capital, and obsolescence, we have a better conceptual foundation for these tasks than we did only a few years ago. Phelps [6], using the same conceptual approach as Solow [7], has estimated the overall rate of return on tangible investment in the U.S. to be about 14 percent in 1954. And even this figure seems conservative in relation to some target rates of return of large industrial corporations reported by Lanzillotti [4].

But whatever the true rates are, they must be compared with appropriate social rates of time preference.

Consider a family of exponential balanced-growth paths sharing a common growth rate; each member of the family has a constant saving ratio, and this ratio differs from path to path. It is also true that each path is characterized by a single technological interest rate, the same for all intervals of time. The theory of the golden rule tells us that the path of highest consumption per capita at every point in time is characterized by a gross saving ratio  $s$  equal to the elasticity of output with respect to capital  $\alpha$  (this is also the share of nonlabor income in GNP if income distribution is governed by marginal productivity). Along the golden rule path the social rate of interest is constant and equal to the rate of increase of the "effective" labor force. This in turn is equal to the natural rate of increase in the labor force plus the annual rate of improvement in labor quality due to technical progress.

If there is no technical improvement, consumption per capita remains constant over time; and along the golden rule path a dollar of per capita consumption saved today will produce a dollar, no more and no less, in



per capita consumption tomorrow. The return on aggregate saving is just enough to keep up with population growth.

This rate of return represents impartiality between generations in this sense: When consumption per capita is the same tomorrow as today, there is no time preference; a dollar of consumption per capita is valued the same whenever it occurs. (See Koopmans [3].)

When there is technical progress, both the real wage and consumption per capita will advance at the annual rate at which labor quality improves, say  $\lambda$ . And along the golden rule path  $\lambda$  will also be the per annum rate of return, in future per capita consumption, on saving today. (A dollar of saving will yield in addition enough new capital to provide for the increment of population.) That is, an increase in per capita consumption of \$1.00 at time  $t$  requires sacrifice of only  $\$e^{-\lambda t}$  at time zero.

It is reasonable to regard this rate of discount, too, as intertemporally impartial. Absence of time preference means that at equal consumption levels society values equally a dollar of future consumption and a dollar of present consumption. But on a path of growing per capita consumption, it is natural that a dollar of future consumption should no longer trade for current consumption at par. To take the rate of improvement in labor quality and in the real wage,  $\lambda$ , as the rate of time preference is to say in effect: saving is justified if and only if it earns more than future consumers will gain anyway through the inexorable progress of technology. Thus if the rate of technical progress is correctly foreseen, this principle meets a common criticism of growth; namely, that there is no reason to save for future generations when technological progress will make them better off anyway. Figure 2 illustrates a social indifference curve between present and future per capita consumption such that there is no time preference when the two are equal, but elsewhere a marginal rate of substitution that exceeds one in the same proportion that future consumption exceeds current consumption.

An economy saving at a constant rate  $s$  lower than  $\alpha$ , the share of capital income in GNP, will be below its golden rule path. Its rate of return on saving will be accordingly higher than the golden rule rate. Indeed the present value of the stream of returns from a dollar of investment, computed at the golden rule rate on the theory that this is an appropriate impartial discount factor free of the taint of time preference, is equal to  $\alpha/s$ . In the United States today the ratio  $\alpha/s$  must exceed 1.5 and may be as high as 2.

For some models it is possible to compute the technological interest rate characteristic of a path with  $\alpha/s$  greater than one; i.e., of a path below the golden rule path. This is, in effect, what Phelps did to arrive at his estimates of the return on investment in the United States, cited above. Consider a model based on a Cobb-Douglas production function

with variable factor proportions both *ex ante* and *ex post*. Let capital elasticity be  $\alpha$  and labor elasticity  $1-\alpha$ ; the natural rate of increase in labor force  $n$ ; constant technical progress expressed as improvement in the quality of labor at rate  $\lambda$ ; a gross saving ratio  $s$ ; depreciation of capital at a constant rate  $\delta$ . The members of this family of growth paths share a rate of growth  $n+\lambda$  in aggregate output, investment, and consumption, and a rate of growth  $\lambda$  in the real wage and in per capita consumption. The rate of interest characteristic of a path is different depending whether technical progress is assumed to be (a) disembodied and affecting all capital old or new, or (b) embodied in new vintage capital only. The expressions for the rate of interest in the two cases are as follows (for their derivation see Appendix):

(a) disembodied technical progress

$$r = \frac{\alpha}{s} (n + \lambda + \delta) - \delta$$

(b) embodied technical progress

$$r = \frac{\alpha}{s} (n + \lambda + \delta) - \delta + \frac{\lambda(1-\alpha)}{s} - \frac{\lambda(1-\alpha)}{\alpha}$$

If, for example,  $n=.015$ ,  $\lambda=.03$ ,  $\delta=.03$ , and  $s=.20$ , then  $r=.095$  in case (a) and  $r=.135$  in case (b). The difference reflects the fact, originally emphasized by Solow [7], that additional saving moves the economy toward a higher path faster in the vintage-capital model and therefore is rewarded sooner with higher consumption.

The evidence is uncertain, and there is a clear need for more refined and reliable estimates of the parameters on which the issue turns. I believe the evidence suggests that policy to accelerate growth, to move the economy to a higher path, would pay. That is, the returns to a higher saving and investment ratio would be positive, if evaluated by a reasonable set of social time preference interest rates. This seems to me the strongest reason for advocating growth policy.

#### REFERENCES

1. Kenneth Arrow, "The Economic Implications of Learning by Doing," *Rev. of Econ. Studies*, June, 1962, pp. 155-73.
2. Edward F. Denison, *The Sources of Economic Growth in the United States and the Alternatives Before Us*, C.E.D., 1962, Chap. 12.
3. T. C. Koopmans, "On the Concept of Optimal Economic Growth," Cowles Foundation Discussion Paper No. 163, 1963, presented at a joint session of the American Economic Association and the Econometric Society on "Intertemporal Economic Theory," in the Boston Meetings, Dec. 1963.
4. Lanzillotti, "Pricing Objectives in Large Companies," *A.E.R.*, Dec., 1958, pp. 921-40.
5. E. S. Phelps, "The Golden Rule of Accumulation," *A.E.R.*, Sept., 1961, pp. 638-42.
6. ———, "The New View of Investment: A Neoclassical Analysis," *Q.J.E.*, Nov., 1962, pp. 548-67.
7. R. M. Solow, "Investment and Technical Progress," in K. J. Arrow, S. Karlin, and P. Suppes (eds.), *Mathematical Methods in the Social Sciences 1959* (Stanford Univ. Press, 1960), pp. 89-104.

8. ———, "Technical Progress, Capital Formation, and Economic Growth," *A.E.R.*, May 1962, pp. 76-86.

## APPENDIX

1. Let  $I(v)$  be gross investment at time (vintage)  $v$ , and let  $\rho(v, t)$  be its marginal productivity at time  $t$ . Then the present value of the stream of returns from investment of one dollar at time  $v$  is

$$\int_v^{\infty} e^{-\int_v^t r(u) du} \rho(v, t) dt.$$

Setting this present value equal to 1 for all  $v$  defines the series  $r(u)$  of instantaneous technological interest rates.

In the models under discussion in the text calendar time does not affect  $\rho(v, t)$ , which can therefore be written as  $\rho(t-v)$ . It follows that  $r(u)$  is a constant, and we may find it from:

$$(1) \quad \int_0^{\infty} e^{-r(t-v)} \rho(t-v) d(t-v) = 1$$

The gross income to capital at time  $t$ , if capital of each vintage is paid its marginal product, is

$$\alpha Q(t) = \int_{-\infty}^t I(v) \rho(v, t) dv = \int_0^{\infty} I(t-v) \rho(t-v) d(t-v)$$

where  $Q(t)$  is gross output summed over all vintages, and  $\alpha$  is capital's share. Now if investment is growing exponentially at rate  $g$ —the rate of growth of output—then  $I(t-v) = I(t)e^{-g(t-v)}$ . Therefore

$$(2) \quad \frac{\alpha Q(t)}{I(t)} = \frac{\alpha}{s} = \int_0^{\infty} e^{-s(t-v)} \rho(t-v) d(t-v)$$

where  $s$  is the saving ratio, constant along the path. The right-hand side will be recognized at the present value of the stream of returns from investment when the discount factor is  $g$  rather than  $r$ . This present value exceeds 1 whenever  $\alpha/s$  exceeds one.

2. The above argument shows that  $r \geq g$  as  $\alpha \geq s$ . It remains to derive the explicit expressions for  $r$  given in the text.

(a) Disembodied progress:

Let  $Q(v, t)$  be the output and  $L(v, t)$  the labor input associated with capital made at time  $v$ .

$$(3) \quad Q(v, t) = A(I(v)e^{-\delta(t-v)})^{\alpha}(L(v, t)e^{\lambda t})^{1-\alpha}$$

The marginal product of capital:

$$(4) \quad \rho(v, t) = \alpha \frac{Q(v, t)}{I(v)} = A\alpha e^{-\alpha\delta(t-v)} e^{\alpha\lambda(1-\alpha)t} \left( \frac{L(v, t)}{I(v)} \right)^{1-\alpha}$$

The marginal product of labor:

$$(5) \quad w(t) = (1 - \alpha) \frac{Q(v, t)}{L(v, t)} = A(1 - \alpha) e^{-\alpha \delta(t-v)} e^{\lambda(1-\alpha)t} \left( \frac{L(v, t)}{I(v)} \right)^{-\alpha}$$

$$w(t)^{-(1-\alpha)/\alpha} = A^{-(1-\alpha)/\alpha} (1 - \alpha)^{-(1-\alpha)/\alpha} e^{(1-\alpha)\delta(t-v)} e^{-\lambda((1-\alpha)^2/\alpha)t} \left( \frac{L(v, t)}{I(v)} \right)^{1-\alpha}$$

$$\rho(v, t) = A^{1/\alpha} \alpha (1 - \alpha)^{+(1-\alpha)/\alpha} e^{-\delta(t-v)} e^{((1-\alpha)/\alpha)\lambda t} w(t)^{-((1-\alpha)/\alpha)}$$

Since the real wage  $w$  grows at rate  $\lambda$ ,

$$\rho(v, t) = A^{1/\alpha} \alpha (1 - \alpha)^{(1-\alpha)/\alpha} e^{-\delta(t-v)} e^{(1-\alpha)/\alpha \lambda t} (w(v) e^{\lambda t})^{-((1-\alpha)/\alpha)}$$

$$\rho(v, t) = A^{1/\alpha} \alpha (1 - \alpha)^{(1-\alpha)/\alpha} e^{-\delta(t-v)} w(v)^{-((1-\alpha)/\alpha)}$$

Thus  $\rho(v, t)$  can be written as  $\rho(t-v)$  and indeed

$$(6) \quad \rho(v, t) = \rho(t-v) = \rho(v, v) e^{-\delta(t-v)} = \rho(0) e^{-\delta(t-v)}$$

To find  $r$  we set  $\int_0^\infty e^{-r(t-v)} \rho(t-v) d(t-v) = 1$

Therefore

$$(7) \quad \rho(0) \int_0^\infty e^{-r(t-v)} e^{-\delta(t-v)} d(t-v) = 1$$

and  $r = \rho(0) - \delta$ . From section 1 we know

$$\rho(0) \int_0^\infty e^{-\rho(0)(t-v)} e^{-\delta(t-v)} d(t-v) = \frac{\alpha}{s}$$

$$(8) \quad \text{Therefore } \rho(0) = \frac{\alpha}{s} (g + \delta)$$

Since  $g = n + \lambda$  we have

$$(9) \quad r = \frac{\alpha}{s} (n + \lambda + \delta) - \delta$$

(b) Embodied progress:

In this case:

$$(10) \quad Q(v, t) = A(I(v) e^{-\delta(t-v)})^\alpha (L(v, t) e^{\lambda v})^{1-\alpha}$$

By reasoning similar to (a) we obtain

$$(11) \quad \rho(v, t) = A^{1/\alpha} \alpha (1 - \alpha)^{(1-\alpha)/\alpha} e^{-\delta(t-v)} e^{((1-\alpha)/\alpha)\lambda v} w(t)^{-((1-\alpha)/\alpha)}$$

$$\rho(v, t) = A^{1/\alpha} \alpha (1 - \alpha)^{(1-\alpha)/\alpha} e^{(-\delta - (1-\alpha)\lambda/\alpha)(t-v)} w(v)^{-((1-\alpha)/\alpha)}$$

Once again  $\rho(v, t)$  can be written as  $\rho(t-v)$ , and

$$\rho(t-v) = \rho(0) e^{-(\delta + (1-\alpha)\lambda/\alpha)(t-v)}$$

The same procedure used in (a) gives:

$$(12) \quad r = \rho(o) - \delta - \frac{(1 - \alpha)\lambda}{\alpha}$$

and

$$(13) \quad \begin{aligned} \rho(o) &= \frac{\alpha}{s} \left( g + \delta + \frac{(1 - \alpha)}{\alpha} \lambda \right) \\ &= \frac{\alpha}{s} (n + \lambda + \delta) + \frac{(1 - \alpha)}{s} \lambda \end{aligned}$$

Therefore

$$(14) \quad r = \frac{\alpha}{s} (n + \lambda + \delta) - \delta + \left( \frac{1 - \alpha}{s} \right) \lambda - \left( \frac{1 - \alpha}{\alpha} \right) \lambda.$$

## DISCUSSION

HARRY G. JOHNSON: A good survey lecture should provide the listener with a set of intellectual calisthenics, beginning with warming-up exercises and gradually building up to feats of reasoning that stretch the mental capacities of the audience. Judged by that standard, Professor Tobin has provided us with a first-class performance: his paper begins with some apparently self-evident commonplaces, which nevertheless effortlessly tidy away many of the conceptual confusions that have plagued current discussions of economic growth as an object of policy; but thereafter his argument becomes increasingly complex, until at the end he is bringing to bear on the question of growth policy the most elegant proposition of modern capital theory: the golden rule of accumulation. As one has come to expect of Professor Tobin, too, his lecture is densely packed with meat. (Perhaps a trifle too much of it comes from the Cowles Foundation sausage-grinders to suit everyone's taste, but a certain amount of institutional growthmanship is allowable on these occasions.) So compact is Professor Tobin's reasoning, indeed, that anyone who has followed it fully on first hearing deserves to be complimented on the extraordinary magnitude of his McLandress Coefficient. I shall not attempt, in these comments, to deal with all of the important issues Professor Tobin has discussed, but shall instead concentrate on a few with respect to which I differ with him in a nontrivial way.

In the main lines of his lecture, Professor Tobin presents and defends what may conveniently be described as the position of the two-G growth man; that is, he favors growth, and he favors government action to increase saving and investment as a means of increasing growth. This position he contrasts with what may be termed the no-G growth position, which he describes as the neutrality of government position and holds up to easy ridicule. I consider the no-G growth position to be a straw man concocted for the occasion—a convenient lecturer's trick, but one hardly conducive to the scientific debate that Professor Tobin calls for. It is, at least, difficult for me to think of a respectable living economist who fits the description, though there are of course some who would espouse the no-G position as part of a broader—and logically consistent—program of reconstructing the economy in the image of classical liberalism. Given the pervasive role of government taxes, expenditures, and other interventions in the economy, government is necessarily an influence on growth as on everything else; neutrality is impossible; and if policy is to be intelligent and not merely blindly conservative, the nature of this influence should be given serious scientific consideration. Where the real basis for debate lies is not between two-G and no-G growthmanship, but between two-G and one-G growthmanship—one-G growthmanship being defined as the position that there are reasons for believing that our institutional structure tends to hamper economic growth, but that some of the restraints originate in the role of government itself, and that in any case macroeconomic measures do not adequately define the proper scope of growth policy.

I shall discuss some of the issues in a moment; but I should first like to point out that the problem of growth policy raises some deep issues of political philosophy, which I do not believe Professor Tobin adequately recognizes. The no-G position, consistently developed, rests on a view of government and the political process as inherently malevolent and untrustworthy. I agree with Tobin in rejecting this view and regarding governmental decisions and actions as having a legitimacy at least equal to that of private decisions in the market. But I do not follow him to the opposite extreme of regarding government as the superego of society, concerned with looking after the best interests of the on-going nation; and consequently I do not regard existing government interventions as having been always designed to serve the social interest. In part, government does seek the social interest; but to an important degree it is a vehicle for the pursuit of private profit maximization by other means than competition in the market. (I cannot, for example, imagine a situation in which citizens in their political capacity tried seriously to stop themselves from privately squandering their wealth.) In large part, also, government is a process of learning from experience—a process subject to a long lag in the correction of errors. I would regard Tobin's view of government as an idealization derived from nineteenth- and twentieth-century historical experience rather than as a reflection of reality; and I would not be too surprised if future social evolution made this conception seem oppressively paternalistic.

I would therefore develop the case for a policy orientated towards growth in a substantially different way than Tobin does. In particular, instead of stopping at the proposition that the role of government in the economy inevitably puts the government in the growth business, I would emphasize the fact that personal and corporate income taxation and commodity taxation introduces a gap between the social and the private returns to accumulation of material and human capital that discriminates against growth. I would also note that in a variety of ways—for example, through minimum wage legislation—social policies can inhibit growth. I would not, incidentally, place the emphasis he does on the effects of taxation on the choice between work and leisure as a factor inhibiting growth. For one thing, if one recognizes as one should the complementarity of leisure with the consumption of commodities, it is not self-evident that taxation generally favors leisure against work. For another, the most casual observation suggests that leisure activity cannot be identified with current consumption. Much of it, such as do-it-yourself home improvement and even more notably education (of which Tobin enthusiastically approves), amounts to accumulation of capital outside the market network. With reference to the imperfections of private capital markets, on the other hand, I would agree with much that Tobin says and commend the caution with which he states his arguments. Some of them, however, require more caution than he exercises, especially those dealing with externalities or alleged externalities. Externalities are intimately interrelated with property rights, and I suspect that in research and development and in the training of labor the present system of property rights is adequate to absorb the externalities more generally than Tobin implies. For example, as Gary Becker

has shown, the cost of acquiring skills that can be transferred from one employer to another will tend to be borne by the employee, via apprenticeship schemes and so forth, and not by the employer. There is also the question, to which Tobin might have given more attention, of how far the government under- or overcompensates for externalities by tax concessions and by bearing part of the expense of investment—a question particularly important in relation to investment in human capital.

As I have already mentioned, the presence and nature of the imperfections inherent in the competitive system and imposed on it by the presence and policies of government lead me to agree with Tobin on the need for policies orientated towards promoting growth. But they do not to my mind lead immediately and exclusively to a macroeconomic policy of increasing saving and investment by fiscal and monetary means. Rather, they suggest the preliminary need for a series of essentially microeconomic reforms of the fiscal system, coupled with policies designed to make the private competitive system produce results more in conformity with the ideal represented by the Pareto optimality of welfare economics. The lines of such policies—development of the insurance principle for pooling risks, social provision of finance for investment in human capital, and so forth—are present but not developed in Tobin's paper. As a case in point, it seems to me rather strange, if Tobin really believes that corporate monopoly is a major obstacle to growth, that he should recommend subsidization of corporate investment by tax concessions or surplus-financed cheap capital rather than the obvious remedy of more determined trust-busting policies. After such reforms, however, it might—and probably would—still turn out that macroeconomic measures for promoting saving and investment were required. In this connection I accept Tobin's view that such measures could be successful, though I would stress more than he does that this depends on an empirical assumption about the response of private saving to increased taxes that is perhaps not so well established as he implies.

I come finally to the question of the payoff from governmentally-promoted increased investment and to the high-powered theorem Tobin uses in this connection—the golden rule of accumulation, or, as it is perhaps better known outside of New Haven, the neo-neoclassical growth theorem. In my own rather old-fashioned judgment, the demonstration that the private rate of return on saving is below the social—and to a significant extent—is sufficient to establish the proposition that increased accumulation would enable the economy to enjoy a potentially superior level of welfare. Reference to the golden rule of accumulation introduces remote and extraneous issues and imposes certain questionable value judgments, without actually solving the problems it raises.

The golden rule merely expresses the conditions, in a dynamic model in which population and technical progress are proceeding at exogenously given rates, for the equivalent of capital saturation in a static model: the maximization of consumption per head. It is a technological relationship, and I consider it somewhat misleading to describe it as "the appropriate social rate of time preference." In so describing it, Tobin is really asserting the proposition



that society should aim at capital saturation; and in applying the concept to the figures, he is merely giving a measure—and the rate of return on investment is not even the most relevant measure—of the extent to which the United States falls short of that happy condition. Whether the golden rule provides an adequate standard for policy is itself debatable. In pursuit of it, a philosopher king might well ordain the abolition of all research and development and the periodic slaughter of most of the innocents, while an individualist citizen might well object to having his savings behavior enslaved to other people's unlicensed fecundity. But if fulfillment of the rule is accepted as a policy objective, implementation of it would demand the immediate sacrifice of all consumption until the capital stock could be built up to saturation level. Since no one, with the possible exception of the Solovians, could take that prescription seriously, the analysis provides no answer whatsoever to the crucial policy question: if the social interest dictates capital saturation, at what rate should society accumulate the necessary capital? It is of little help to know where one should be going, if one is not told when one ought to arrive.

In conclusion, I should like again to congratulate Professor Tobin on an excellent lecture. He has attempted the extraordinarily difficult task of bringing order to a new field of economic policy, in which the issues range from the naivest forms of slogan-mongering ("get out and grow") to the most elegant and sophisticated of formal theorizing ("follow the golden rule"); and he has succeeded brilliantly in presenting his views of those issues succinctly and intelligibly.

HERBERT STEIN: I am pleased to be able to join his colleagues in honoring James Tobin. We honor him for the many fruitful ideas he has contributed to our thinking and not less for the example he has provided of the proper demeanor of an economist. He has participated modestly and responsibly in the discussion of economic issues without succumbing to the lure of interpersonal, interschool, or interparty warfare. Professor Tobin's paper embodies these qualities, and I am sure I speak for all present in congratulating him on it.

This paper will do much to place economic growth in its proper position among the objectives of government policy. As the paper says: "Let growth be something it is possible to oppose as well as to favor, depending on judgments of social priorities and opportunities." Particularly, the clear distinction drawn here between growth and full employment will, if taken to heart, dispel much misunderstanding.

While the paper gratifies me it does not entirely satisfy me. At the risk of seeming greedy, I would like to raise some questions in the hope that subsequent writing of Professor Tobin and others will provide answers.

The paper presents "a typical social indifference curve between present and future per capita consumption such that there is no time preference when the two are equal, but elsewhere a marginal rate of substitution that exceeds one in the same proportion that future consumption exceeds current consumption."

Why should we think that the appropriate indifference curve has this prop-

erty? I would think that there are many possible indifference curves that are impartial as between generations but have different marginal rates of substitution when the incomes are unequal.

Professor Tobin is suitably modest in his claims for this indifference curve. He calls it "a typical social indifference curve," by which I suppose he means typical of the curves economists draw for pedagogical reasons and not based on any evidence. Again, he refers to "the theory that this is an appropriate discount factor." But in the end the conclusion that policy to accelerate growth would pay seems to depend on the slope of this hypothetical curve.

Suppose that we could, by revelation or otherwise, learn what the appropriate social impartial discount factor free of the taint of time preference is, would we be sure that this justifies a higher rate of saving than occurs as a result of private decisions tainted by time preference? The private decisions are tainted in various ways that tend to increase saving.

First, we are not impartial between generations. The next generation is our children and we are a child-oriented society. There is an old story about the immigrant who never in his life got to eat the white meat of chicken. When he was a child in the old country the parents got the white meat and when he became a parent in America the children got the white meat. We may all value white meat tomorrow less than white meat today, but some of us value white meat for our children more than white meat for ourselves.

Second, I suspect, though I have taken no census, that people in general tend to underestimate the rate at which incomes will rise through time and therefore overestimate the need for saving.

Third, people save for protection against private risks which would not require saving by the society as a whole.

Fourth, apparently a large sector of the population is driven by what a colleague of Professor Tobin has called "the Puritan ethic," which embodies a strong compulsion to save.

However, let us assume that we have discovered that we are saving too little. Let us assume further that we have persuaded everyone to accept the moral imperative: take what thou hast and give to the rich, that is to the future, so that they may be richer still. I shall not stop to ask what arguments could be used to persuade someone who did not find this idea spontaneously appealing. I would like to ask how the government gets into this picture, since Professor Tobin's paper is about growth as an objective of government policy.

Why should individuals require the help of government in order to carry out temporally impartial savings decisions if that is what individuals want to do? Ample savings media exist without further action by government. The effectiveness of one individual's saving does not depend on whether other people are also saving. There is no critical mass that the savings must reach. It is not necessary that everyone should save in some particular way, which has to be specified by a central authority.

I suppose that government enters the picture if an effective majority of the population desires to save at the socially appropriate rate but some significant minority does not. Then government becomes the instrument by

which the majority forces the minority to save more than it wishes. I do not think we can brand this as illegitimate a priori. Government is an instrument of force, and we use it properly for many important purposes. But the use of force deprives some people of some part of their freedom. This is a cost and it should be minimized, or not incurred except for highly important gains. If we take the position that every departure from any optimum calls for government action, that not a sparrow falleth without becoming a concern of government, I fear we shall have more government than we can manage and less freedom than we would like.

In a special way, Professor Tobin's paper recognizes this. I refer to his few comments on population policy. If we are concerned, as Professor Tobin suggests we should be, with the growth of per capita incomes, it would seem natural that we should be concerned with the growth of population. But he accepts the fact that we do not have, and should not have, a government population policy. I think this is because he recognizes that the private freedom involved here is too important to be impaired for the sake of optimizing the rate of growth of per capita consumption.

Does not the same kind of question arise, even though on a different emotional scale, with respect to individual's freedom to use their own incomes? Don't we have to ask not only whether saving and investment are below the social optimum but also whether the deficiency is so great and the consequence so serious as to justify government action?

This leads me to Professor Tobin's discussion of government neutrality towards growth. He demonstrates convincingly that government cannot be neutral with respect to growth, but inevitably affects growth in a variety of ways. If I have used the word "neutral" in this connection in the past, I recant. I would offer as a substitute for the word "neutral" the word "indifferent." Surely the government can be indifferent about economic growth. We would not expect the government to have a policy about every consequence of its actions. The pattern of government expenditures affects the distribution of the population among states and localities. But this does not require the government to have a policy about the regional distribution of the population, and we do not, except in extreme cases where there is distress. Something does not become an objective of government policy simply by being a consequence of government policy.

I do not want to run the notion of indifference into the ground. I only offer it as a feasible alternative to growth policy—more feasible than neutrality. There may be cases in which government action impairs growth and serves no other useful purpose. Such cases would call for correction. But we would expect these cases to be few and quantitatively unimportant. The big government actions that affect growth also have other objectives and effects. The import of a policy of indifference is that these actions should be determined by consideration of objectives other than speeding up growth—primarily and in general.

In the end, as Professor Tobin said at the beginning, the question is whether more growth is worth its cost. He has computed the cost in terms of present consumption foregone, compared it with the discounted value of the

future consumption gained, and concluded that there would be a net gain. Whether such a computation can be done with any objectivity, especially in view of the difference in income levels between present and future generations, is doubtful. Whether, however it is done, the result would come out in the direction indicated, is at least uncertain. But if I remain skeptical, it is not primarily for these reasons. It is primarily because of the costs not included in this calculation. These are the costs of the measures that government may have to take in order to make us save substantially more. Among these possible costs are reduction in our freedom to dispose of our own incomes, growth in the role of government as investor, impairment of the equity of the tax system, reduction of important government expenditures that do not qualify as investment, and diversion of the limited problem-solving capacities of government from other critical issues. These are the risks that make me hesitate to raise the further enrichment of our grandchildren into the top ranks of government policy objectives.

If any grandchildren of mine should ever read these words, I hope they will not think that I disregard their interests. I expect that they will be richer than we are, and hope that they will be even richer than I expect them to be. But to make them richer still is not, in my opinion, among the most valuable things we can do for them. It is much more important to pass on to them a world of reasonable security, in which they have substantial freedom to manage their own affairs and dispose of their own incomes, in which the relatively poor among them are relatively less poor, and in which the Indians and Guatemalans and Ghanaians are much less poor relative to the Americans. If we can do this, our grandchildren will have no cause to think bad of us. Perhaps we can achieve this and still accelerate future growth. But our energies and intelligence, our capacities for leadership and followership, our ability to govern ourselves, our stock of objectivity and altruism are all limited—more limited than our national income. I fear that if we set ourselves too many high-priority goals we shall achieve none of them. The best is the enemy of the good. May we not have to forego the effort to achieve the best of all possible worlds in order to increase the chance of achieving the merely good?

# THE THEORY OF MONOPOLISTIC COMPETITION AFTER THIRTY YEARS

## THE IMPACT ON INDUSTRIAL ORGANIZATION

By JOE S. BAIN

*University of California, Berkeley*

The *Theory of Monopolistic Competition* appeared in 1933, and by the later 1930's a new approach to the empirical investigation of industries and markets emerged, the pursuit of which resulted in the development of a field of study in economics early labeled and still known as "industrial organization."

The new field differed enough from earlier related fields to be recognized as substantially novel, although it did have progenitors or parents. If a simple biological analogy were applicable, we might say that Industrial Organization was foaled out of Trust Problems by Price Theory, but in the world of ideas multiple parentage is common. In this case, I am satisfied to name price theory as the principal sire, but on the side of the dam we must look to a complex of what were frequently labeled as "institutional approaches" to the industrial economy, including that of marketing as well as that of trust problems, including descriptive studies of price and profits, and including the pioneering work of Berle and Means. However this may be, the offspring was distinctive, exhibiting as it grew the capacities to erase artificial barriers between theoretical and institutional studies of markets, to blend them in organized empirical investigations which had definite theoretical orientations and were related to full-blown sets of hypotheses, and even to generate additions to or elaborations of the theory which was its parent.

Can we be a little more specific about the parentage? I believe so, if we look at the offspring. Although scholars working in and around the field have studied and written about many things, students of industrial organization proper not only have emphasized an external view of the market conduct and market performance of groups of firms (congruent with the emphasis of price theory and differing from the internal or industrial-management view), but also have placed a predominant emphasis on the study of the market structures of industries, on interindustry differences in structure, and on associations between market structure and market performance. Their disposition to study and classify market structures clearly reflects the belief that

such structures do differ significantly, in the sense that structural differences are likely to be associated with significant differences in performance. Their search for the existence of such associations can be regarded generally as attempts to confirm or disconfirm theoretical hypotheses predicting such associations, to discover needed elaborations in the system of hypotheses, or to provide empirical answers to questions concerning associations of market structures to performance which the theory of the moment asks but does not answer.

This major emphasis in the empirical study of industry, the principal concepts on which such study depends, and indeed the field of industrial organization as we now know it, all find their primary theoretical origin in Chamberlin's *Theory of Monopolistic Competition*. For it was this work which simultaneously did two very important things. It advanced expressly—really for the first time—the major and crucial theoretical construct in which an economy of enterprises was viewed as being made up of industries having a variety of distinctly different market structures, with market conduct and performance tending to differ significantly with differences in structure. And it implemented this construct by developing an uncomplicated but actually quite sophisticated classification of market structures. This classification reflected a remarkable insight into what in the world of markets was empirically relevant to price theory, was the evident parent of substantially all more elaborate market classifications which have been subsequently suggested, and provided the basic skeleton for all of them which have had appreciable merit. Suddenly, therefore, economists were presented with a sophisticated theory of markets in lieu of a theory of perfect competition with dangling appendages concerning monopoly and economic warfare.

Like many theoretical contributions of revolutionary character, Chamberlin's seems at retrospective first glance to have been quite simple, or, let us say, artistically simple. In reformulating price theory, the basic novel things that Chamberlin did were to introduce the notion that groups of rivalrous firms often sell differentiated products and explore its implications, to resurrect and reformulate a long-lost theory of duopoly, and to expand it into a major novelty, a theory of oligopoly. I would regard the general contributions embodied in his analysis of oligopoly as overwhelmingly important, and not really so very simple. For it at once suggested the need for a theory built to fit a substantial proportion of all industries or markets in the real economy, which would fill a gaping hole in pre-existing price theory; developed *de novo* an original pure theory of oligopoly which was both sensible and novel; and suggested the major principles according to which the broad oligopoly category in a general market classification might be

broken into a number of subcategories with significantly different structural characteristics. In the twelve pages which comprise his three sections of "mutual dependence recognized," "the effect of uncertainty," and "the small group: oligopoly plus product differentiation," Chamberlin revised modern price theory drastically and permanently. My emphasis on the importance of this aspect of his contribution is not intended to minimize in any way the correlative importance of his imaginative exploration of the broader implications for price theory of recognizing formally the existence of product differentiation, product changes, and selling costs.

Chamberlin's basic construct and his implementation of it were extremely important because of their obviously great empirical relevance to the actual economy of business enterprises, which, once one thought about it, was clearly composed of a mixture of pure and heterogeneous oligopolies of several distinct structural subvarieties, atomistic markets with product differentiation, and some markets with pure competition. His contribution, therefore, was not "timely" (in the sense that Keynes's *General Theory* was timely); it was a long way overdue, and had *ex post* applicability running back to 1890 and before as well as current and future applicability. It has occurred to me, amidst subsequent and perhaps legitimate quibbles over such matters as how much a seller's species demand curve in monopolistic competition really slopes, that the utterly unshakeable importance of Chamberlin's major construct and his development of it is sometimes forgotten.

This, then, was the major impact of the *Theory of Monopolistic Competition* on the development of the field of industrial organization—its development and general design were really suggested by Chamberlin's major work. His initial market classification, substantially adequate, has been fruitfully elaborated. By providing the major outlines of a theory of oligopoly, he identified a crucial area and direction for further theorizing and for empirical research that have ensued. And he supplied the crucial construct and its basic pattern of implementation. Further, his identification of dimensions of market performance in addition to those involving price, production cost, and output—that is, selling cost and product quality and variety—suggested an expanded range of matters deserving empirical study within the sort of theoretical framework he had established. Further yet, he suggested several fruitful hypotheses concerning market conduct in imperfectly collusive oligopolies.

The subsequent development of the field of industrial organization of course depended on one who would grasp the broad implications of the Chamberlin construct and classification of market structures,

elaborate the latter, devise means of organizing available data for the testing of hypotheses concerning structure-performance links, and encourage theoretically-oriented empirical research of both the individual-industry and cross-sectional types. (Not the least of the contributions at this stage was the proposition that, since the new sort of theory of markets suggested structure-performance links, it might be confirmed or disconfirmed by comparing observable structural characteristics of markets with observable performance, without ever learning much about essentially unobservable sellers' species demand curves.) The party principally responsible for all of this is of course Edward Mason. A very large amount of progress in the field of industrial organization to date has resulted from running down leads that he initially suggested during the 1930's.

There are of course some who hold that price theory can do nicely without oligopoly and product differentiation (and thus without the *Theory of Monopolistic Competition*), and implicitly that a study of industrial organization based on a sophisticated theory of markets is not necessary. A theory of competitive price will be sufficient, or the old competition-monopoly dichotomy will do. This could follow from the proposition that the economist in a bewildering world should be content if he can hit a dinosaur with a scattergun at ten paces. Some difference in aspiration levels seems to be involved here, or some disagreement about the character of our surroundings. Chamberlin's work essentially suggested that we take a close look at our surroundings, check the calendar, and use more adequate weaponry. I am inclined to agree.

I have so far emphasized that the impact of Chamberlin's work on the development of the field of industrial organization was crucial. In so doing, I do not wish to discount the considerable correlative impact of other work in price theory which antedated or was contemporary with his. First, of course, everyone owes a debt to Alfred Marshall. Second, the formal theory of the firm and of the perfectly competitive market was greatly developed during the 1920's, and Chamberlin's work depended on it, as well as being influenced by "off-the-beaten-track" works like Zeuthen's *Economic Warfare*. In addition, the work of a number of writers, including E. A. G. Robinson and P. Sargant Florence on scale economies for the plant and firm, must be recognized as having independent importance.

It would hardly be correct, however, to say that Chamberlin merely put an obvious capstone on a pre-existing edifice. It had not been obvious (stray footnotes and isolated theorems to the contrary notwithstanding), and in any event Chamberlin rebuilt the edifice to a considerable extent.



An important independent impact (the effectiveness of which depended in considerable part on the availability of the Chamberlinian construct) was made by the work of Joan Robinson on monopsony and on price discrimination (as by later work on bilateral monopoly), which suggested elaborations of the theory of markets which have considerable empirical relevance. As to Mrs. Robinson's work in general, it may be argued that some of what Chamberlin made express is implicit in her work. The term implicit is used advisedly, since it is easy to find suggestions of the major Chamberlinian construct and its implementation in her work only after reading Chamberlin; lacking this, it is easier to find the "box of tools for the analytical economist" (to use Mrs. Robinson's words), the provision of which was the asserted purpose of her excellent work.

Acknowledging the importance of the theoretical works mentioned and of others written prior to and since 1933, it must yet be recognized that the *Theory of Monopolistic Competition*, by enunciating in express form a novel theoretical construct and implementing it to provide the essentials of an empirically relevant and sophisticated theory of markets, was *the* major contribution to price theory which set in motion the systematic development of the modern field of industrial organization, and thus a blending of abstract price theory and empirical research which aims to be scientific rather than philosophical.

## THE IMPACT ON GENERAL THEORY

By ROBERT L. BISHOP

*Massachusetts Institute of Technology*

The character and importance of the impact that the theories of monopolistic and imperfect competition have had on general theory obviously depend on whose general theory it is. As I judge the consensus of economists, Chamberlin's *Theory of Monopolistic Competition* and Mrs. Robinson's *Economics of Imperfect Competition* are acknowledged to have touched off, in 1933, a theoretical revolution whose relative importance in the microeconomic area was comparable to that of Keynesian analysis in macroeconomics. But there are also some significant minorities in the profession who would minimize or reject completely the positive value of the Chamberlinian and Robinsonian doctrines; so, for anyone like myself that happens to hold these theories in high esteem, it is important to test that judgment against the counterarguments of the opposition. Accordingly, I shall divide my remarks between, first, a brief characterization of the merits and positive significance of Chamberlin's and Robinson's contributions and, second, a consideration of the nature and worth of some of the dissenting views.

Let me begin on a personal note. I was privileged to begin my own study of economics, as a freshman at Harvard, just before the Chamberlin-Robinson revolution—in the very year that their pioneering works were published. I confess that I was a better student of the received doctrines that were set before me than I was an observer of reality; and, among other things, I absorbed quickly the notion that “supply and demand” were the key to an understanding of how any market operates, including the occasional monopolistic one that was subject to a control and restriction of supply. Then, sometime during that first year, I heard someone talking about a distinction between buyers' markets and sellers' markets, according as buyers found it easy to buy or sellers to sell. My own instant reaction was to classify the speaker as an economic illiterate; for it seemed perfectly apparent that, in any “competitive” market, the equilibrium price had to be at the level where demand and supply were equal, so that every buyer could readily buy, and every seller could no less readily sell, whatever respective quantities they pleased. When I subsequently learned that the speaker was indeed talking about relevant economic phenomena, not only did I make a belated mental apology to him for my similarly

unspoken slur on his intelligence, but I also felt that I had acquired a justifiable grievance against the analytical tradition that had trapped me into placing my faith in the universality of the law of supply and demand. By this time, of course, I had become a student of monopolistic competition, having found readily convincing the judgment that the demands relevant for most sellers' price decisions were significantly less than perfectly elastic. This was a conclusion that seemed to me then, and still seems to me now, essentially inescapable as soon as the equilibrium of the individual firm was subjected to a more explicit and cogent analysis than, say, in Marshall's fuzzy device of a "representative firm."

Now I do not mean to suggest for a minute that my betters among pre-1933 economists were all equally and uniformly guilty of the same overliteralness in their understanding and analytical use of the law of supply and demand than I was as a freshman. I do believe, however, that they were frequently victims of the state of affairs in which their only explicit theory of competition involved what is now called pure or perfect competition. Furthermore, having only that theory, they were also subject to the persistent analytical bias of a wishful thinking that the world did indeed conform well enough, for the most part, to the conditions necessary for a relevant application of the only theory at their command. Nor has this bias altogether disappeared; it is still very much alive, for example, in such citadels as the University of Chicago. As always, it takes a theory to lick a theory. Accordingly, whether we call the phenomenon "monopolistic competition" as in this Cambridge or "imperfect competition" as in the other one, the intensive theoretical treatments of Chamberlin and Mrs. Robinson were essential in order to force upon the attention of the economics profession due regard for the fact that most product markets differ significantly from the traditional models of either simple monopoly or pure competition.

I can afford to sketch only briefly, and with a few selected examples, some of the ways in which these doctrines have made a difference in economic theory and in the views that most of us entertain as to the economic world around us. Example: as late as 1932, even as excellent an economist as J. R. Hicks was able to write a *Theory of Wages* in which he was still telling us that wages are typically equal to the value of labor's marginal product. In the commentary that he has recently published on that early work, he has himself acknowledged to some extent the seriousness of the defects attributable to his having so consistently and uncritically assumed perfect competition in both product and factor markets.<sup>1</sup>

<sup>1</sup> J. R. Hicks, *The Theory of Wages* (2nd ed., 1963), p. 310.

A few examples concerning Marshall's economic theory are especially appropriate; for he was probably without peer in the delicate art of not letting his inadequate theory get too much in the way of his sensible view of reality. As to retail trade, for example, he was usually careful to mention that the competitive theory he was talking about did not apply.<sup>2</sup> Now, of course, we have a theoretical framework into which an analysis of many different kinds of retail trade can be fitted. Or consider the famous category of the decreasing-cost industry, whose existence Marshall had to explain by means of a very confusing set of comments about the internal and external economies reaped by the representative firm when demand for the industry's product went up, despite the fact (which he uneasily recognized) that unexhausted internal economies of scale are incompatible with purely competitive equilibrium. With product differentiation or oligopoly (or both), the possibility that increased demand may lead to lower prices becomes eminently comprehensible and plausible, without recourse to any mystique about an especially dubious type of external economy. Finally, consider the anomalousness of Marshall's view that price typically equals marginal cost except when demand is too low for the firm to cover its full costs, whereupon misgivings about "spoiling the market" are dragged in to explain why the firm does not suffer the even greater loss that would be implied if it were to continue to equate price to marginal cost (at least down to minimum average variable cost). This conception is revealed in all its naked implausibility by Frisch's graphical interpretation of Marshall's verbal exposition.<sup>3</sup>

As soon as we recognize that there are comparatively few markets in which "demand and supply" operate in the essentially symmetrical way that is the essence of that famous law, many things become analytically clear for the first time. Thus, in product markets where buyers are passive and prices are unilaterally administered by sellers, the normal condition of price in excess of marginal cost explains why it is the sellers rather than the buyers who are eager for more transactions. Not only does this open the door to such phenomena as advertising and other forms of non-price competition, but it also helps explain why we are a race of eager sellers and coy buyers, with purchasing agents getting the Christmas presents from the salesmen rather than the other way around, and with "salesmanship" a familiar concept while "purchasemanship" does not even appear in the dictionary. We find here also, I am convinced, part of the explanation—even though not the whole—of the persistent producer-orientation of so many of the world's mercantilist policies, both ancient and modern, international and domestic.

<sup>2</sup> A. Marshall, *Principles of Economics* (8th ed., 1920), p. 328.

<sup>3</sup> R. Frisch, "Alfred Marshall's Theory of Value," *Q.J.E.*, 1950, pp. 495-524.

Not the least of the contributions of these doctrines, in my opinion, is the improved sense that they make out of the major propositions that have emerged from Keynesian macroeconomic analysis. Keynes himself, of course, was still sufficiently Marshallian when he wrote the *General Theory* so that he based his theory on an assumption of universal pure competition, but with the anomalous extra feature of a perfectly elastic labor supply up to the full employment quantity, as based on a peculiarly strong and implausible "money illusion" on the part of the labor suppliers. This was Keynes's way of explaining how an underemployment equilibrium was possible without a persistent price deflation. In contrast to that rationalization, how much more natural and plausible it is to explain the necessary price and wage rigidities or "stickiness" with reference to the fact that the world is not predominantly purely competitive.

Similarly, it is only with reference to these doctrines that we can make analytical sense out of the more modern dilemma that really full employment is very hard to achieve without at least a creeping inflation. The difficulty is not at all that some prices and wage rates rise for the good competitive reason that otherwise there would be an excess of demand over supply, but rather that other prices not only fail to fall but are even capable of being increased when the opposite condition prevails. This asymmetry, whereby many prices and wage rates go up much more readily than they go down (save only in the deepest depression), is again traceable to the basic facts that (1) our price-determining machinery is preponderantly seller-dominated, and (2) the normal condition of such markets is that supply exceeds demand—in the relevant sense that sellers (whether of goods or labor) are willing and eager to sell more than they can actually sell at their current prices. To investigate the conflicting claims of demand-pull and cost-push theories of inflation without due regard for these considerations, as some scholars still do, is to omit a most important ingredient of the analysis.<sup>4</sup>

Widely accepted though the theories of monopolistic or imperfect competition have been, there are still some determined pockets of resistance. I turn, accordingly, to an examination of some of the views entertained in those quarters. The various misgivings can be disentangled to some extent, I think, under such headings as (1) ideological, (2) methodological, and (3) factual.

Even though ideological precommitment is a poor influence in shaping anyone's evaluation of a new theory, such ideological contamination has played a not inconsiderable role in just such areas as the

<sup>4</sup>For example, R. T. Selden, "Cost-Push versus Demand-Pull Inflation, 1955-57," *J.P.E.*, Feb., 1959, pp. 1-20.

present one. As always, ideology cuts at least two ways. Thus, in the atmosphere of the depressed 1930's, so conducive to the deepest disillusionment with the current performance of capitalist free enterprise, it was not surprising that many young Turks—and some old ones as well—were strongly predisposed to find in the theories of monopolistic or imperfect competition, not only a legitimate analytical framework, but also a device for dramatizing in an exaggerated way at least some of the sins of capitalism. Mrs. Robinson herself, I am told, has testified that the central motivation for her own book was to demonstrate the pervasiveness of the exploitation of labor; and her bias is further apparent in her failure to give a corresponding attention to monopoly power in labor markets. It is sometimes inadequately appreciated that, no matter how much labor a great corporation may hire, it exerts no monopsony power at all unless it is unable to get as much labor as desired at the current wage rate; conversely, if there is an excess of job seekers at that wage, the predominating influence in the situation is monopolistic, not monopsonistic.

On the other side, the central ideological resistance to a due acknowledgement of the widespread relevance of monopolistic competition comes from those theorists who have a strong emotional stake in an absolute and uncompromising defense of the optimality of unregulated markets. This is illustrated, for example, in the writings of Mises and Hayek. To Mises, imperfect or monopolistic competition is "mythology" (*Human Action*, 1949, p. 378). His argument is a curious one, however, for he concedes the existence of all of the basic phenomena with which the theory of monopolistic competition is concerned, including product differentiation and oligopoly (e.g., pp. 354-61). Indeed, he even concedes that "almost all consumers' goods are included in the class of monopolized goods" (p. 357). But, even more curiously, he also argues that it does not follow by any means that monopolists necessarily, or even presumptively, charge monopoly prices (e.g., p. 356). Now it should not be thought that Mises is talking about the *curiosum* of a "monopolist" who faces a perfectly elastic demand; for he is ludicrously confused about this. Thus, a bit farther on (pp. 381-82), in speaking of the various ways in which the individual consumer may react to monopoly prices, he says of one of the possibilities: "The consumer restricts his purchase of the monopolized commodity to such an extent that he spends less for it than he would have spent under the competitive price; he buys with the money thus saved goods which he would not have bought otherwise. (If all people were to react in this way, the seller would harm his interests by substituting a higher price for the competitive price; no monopoly price could emerge. . . .)" The cream of this little jest,

of course, is that a monopolist with positive marginal cost maximizes his profit only where marginal revenue is also positive, or where demand exhibits the relative elasticity that Mises' case implies.

But if Mises' own discussion founders in this type of futility, he also invokes the aid of a much more careful and competent analyst when he cites (p. 278, n.) as "a refutation of the fashionable doctrines of imperfect and of monopolistic competition" two essays by Hayek in his *Individualism and Economic Order* (1948; pp. 92-118). Actually, however, Hayek's unhappiness with the concept of an "imperfect" market really rests primarily on his misgivings as to the unreal character of a "perfect" one and on what he describes as "the so-called theory of 'perfect competition'" (p. 92). Thus, his quarrel would seem to me to be less with Chamberlin and Mrs. Robinson than with such theorists as Friedman and Stigler, who defend the application of the perfectly competitive model in many situations where others of us regard it as very dubiously applicable.

Specifically, Hayek argues "that what the theory of perfect competition discusses has little claim to be called 'competition' at all and that its conclusions are of little use as guides to policy" (p. 92). This reflects a methodological complaint that "the modern theory of competition deals almost exclusively with a state of what is called 'competitive equilibrium' in which it is assumed that the data for the different individuals are fully adjusted to each other, while the problem which requires explanation is the nature of the process by which the data are thus adjusted" (p. 94). In reply, it seems to me that no competent theorist regards static-equilibrium theory as the last word of economic analysis. Moreover, part of our interest in static equilibria concerns their stability conditions, from which the more general comparative-statics theorems are generated; and, beyond that, everyone stands ready to move on to explicitly dynamic analysis whenever necessary. But to argue, as Hayek seems to, that dynamic analysis can be wholly divorced from concepts of equilibrium seems to me anything but fruitful. In other words, even the dynamics of a market remain quite incomprehensible except as we can analyze them as movements toward or away from some equilibrium position, even when the data are changing so rapidly that at least the longer-run types of equilibria are never actually attained. In short, Hayek's discussion is less a "refutation" of what others call imperfections of competition—like Mises, he also willingly acknowledges the phenomena and balks only at what they are called—than it is an obscurantist effort to undermine all of the standard techniques of economic analysis.

On the other hand, Hayek is too sensible a man to hold consistently

to his rejection of equilibrium analysis. Thus, his methodology does not prevent him from comparing "an 'imperfect' market . . . with a relatively 'perfect' market as that of, say, grain" (p. 102), despite his previous argument that competition has to be not only dynamic but personal (pp. 96-97). On the latter point, he cites as "remarkable" Stigler's eminently perceptive statement that "economic relationships are never perfectly competitive if they involve any personal relationships between economic units" (in the sense, of course, that when demand equals supply in a perfectly competitive market, it makes no difference to any buyer or seller as to the identity of the other party or parties with whom he deals, and also in the sense that no buyer or seller is thereby frustrating any other person in his desire to make a similar purchase or sale at the competitive equilibrium price). My final comment is that, if Hayek finds such impersonal equilibria unrealistic, he is more than welcome to join the ranks of the converts to the theory of monopolistic competition. Nor, if he were to accept this invitation, should he fear that he would thereby be forced to give up what he cites as the "practical lesson of all this, . . . that we should worry much less about whether competition in a given case is perfect and worry much more whether there is competition at all" (p. 105). Surely, many of his presumed opponents agree wholeheartedly with that sentiment, and also with his further remark that "much more serious than the fact that prices may not correspond to marginal cost is the fact that, with an entrenched monopoly, costs are likely to be higher than necessary." To this I would only add that so may they be when oligopolists invite excess capacity with high-price policies and deliberately create excess capacity to discourage further entry, or when monopolistic competitors of any kind inflate their costs with certain types of sales-effort expenditures as an alternative to cutting prices. But even if Hayek rejects this invitation, the rest of us must at least be pardoned if we conclude that his discussion constitutes a pale and unconvincing "refutation" of the doctrines in question.

A quite different motive for either ignoring or minimizing market imperfections is rooted in the theorist's methodological bias in favor of the assumption of pure competition, because of its analytically tractable features. Sometimes this assumption is made in a properly apologetic way, as by Hicks in *Value and Capital*. Still, this does not keep him from consoling himself with the comforting thought: "I do not myself believe that the more important results of this work are much damaged by this omission [of imperfectly competitive influences]" (p. 7). And later, when he acknowledges that certain considerations do indeed have "to be met by sacrificing the assumption of perfect competition," he adds unhappily, "yet it has to be recog-



nized that a general abandonment of the assumption of perfect competition, a universal adoption of the assumption of monopoly, must have very destructive consequences for economic theory" (p. 83). In other words, he feels that it is better to have a neat, inapplicable theory than to face up to the difficulties of formulating some relevant ones.

Where Hicks is at least frankly apologetic and defensive, however, others such as Friedman assume an aggressive initiative, even to the extent of finding positive merit in the failures of the purely competitive model to conform better than it does to the facts of a great variety of situations. If I understand him aright, Friedman imposes on himself the heroic act of self-denial of choosing to work with only two abstract models (*Essays in Positive Economics*, pp. 35ff.). These involve: (1) a "competitive" firm, where "the demand curve for its output is infinitely elastic with respect to its own price for some price and all outputs, given the prices charged by all other firms"; and (2) a "monopolistic" firm, where "the demand curve for its output is not infinitely elastic at some price for all outputs." In a footnote, however, he breaks the latter category down into two types: (1) "the monopolistic firm proper, if the demand curve is nowhere infinitely elastic (except possibly at zero output)," and (2) "the oligopolistic firm, if the demand curve for its output is infinitely elastic at some price for some but not all outputs."

As to this weird oligopoly concept, it may first be noticed that oligopolistic suppliers of merely close (but not perfect) substitute products are not oligopolists at all to Friedman, but just monopolists. Thus, there is apparently no competition worth analyzing at all between Ford and General Motors. Second, there is not even a clear distinction between competition and oligopoly in Friedman's scheme; for, whether the suppliers of perfect substitutes are two or two thousand, the demand confronting any one of them (with other prices constant, as Friedman specifies) is perfectly elastic only up to the total quantity demanded in the market as a whole at the common price. On the other hand, Friedman never puts his oligopolistic model to any analytical use anyway; so it does not really constitute a third model as an addition to the two basic ones.

Having only two such models at his disposal, then, he naturally needs a good deal of leeway in the discretion that he allows himself as to their use. Thus, as he makes perfectly clear in an example relating to the cigarette industry (pp. 37-38), either model is relevantly applicable depending on the problem at hand. First, however, no guidance whatever is offered as to the principles that might determine which of the two models is relevant, in advance of applying one or the

other. Second, there is obviously a lot of room here for self-deception as to whether either model really applies at all satisfactorily. Thus Friedman says: "Suppose the problem is to determine the effect on retail prices of cigarettes of an increase, expected to be permanent, in the federal cigarette tax. I venture to predict that broadly correct results will be obtained by treating cigarette firms as if they were producing an identical product and were in perfect competition." There are some problems, however, as to how this conjecture might be tested. Thus, depending perhaps on the period allowed for the industry's adjustment, the gross price paid by consumers will presumably rise; but any such reaction can be tautologically rationalized after the fact with reference to either competitive or monopolistic theory, whether the price rises by an amount greater than, equal to, or less than the tax. Clearly, if competitive theory is to be relevantly applicable, Friedman must first find not only a supply curve for the cigarette industry, but the whole collection of successively relevant short-run and long-run supply curves; and he must then find that the cigarette industry does indeed behave "as if" it were constantly equating demand and supply. Now, actually, there is a test that can at least disprove what Friedman ventures to predict. In an industry that really is purely competitive, with the basic demand and cost data constant, a tax may cause the price to rise little, if at all, in a very short run; then it causes the price to rise more in each successively longer short run; and finally it rises most in the long run. But monopolists and oligopolists, in their wisdom as dynamic profit maximizers, often respond to a tax with an immediate, once-and-for-all price increase—quite unlike pure competitors. My own ventured prediction is that cigarette manufacturers do indeed respond in this oligopolistic manner, perhaps even rescinding part of the price increase in the longer run if a reduction of factor prices is induced by the curtailment of factor demand.

The same sort of determination to divide both reality and its economic analysis into the two neat boxes of competition and simple monopoly is also illustrated in Stigler's work. Thus, in one of his *Five Lectures on Economic Problems* (1950), he concludes that the theory of monopolistic competition is a "failure" (p. 22) because it embraces too many diverse conditions for any single, neat theory to cope with in a fruitful way. Then, in the last lecture, he gallantly undertakes to classify as best he can the industries of the United States into the two categories of "competition" and "monopoly" (with a monopolistic subcategory of "compulsory cartels"). "A competitive situation," he says (p. 48), "is characterized by price or product competition of the individual firm, leading to approximate equality of price and marginal cost in the long run." This is a constructive start, but it has two

weaknesses: (1) the standard of "approximate equality" is not clear (should it be no more than a 2 percent disparity, for example, or would 10 percent be acceptable too?), and (2) what is the meaning and the justification of limiting the prescription to the long run? As to the first point, it is also important to appreciate that, with production divided into various vertical stages, the magnitude of price minus marginal cost is more relevantly compared with value added than with price. As to the reference to the long run, if this means only that price may temporarily exceed marginal cost by an appreciable percentage, for example in the early stages of major innovative change, then well and good; but if it means, as I fear it does, that price may acceptably be only slightly higher than long-run marginal cost at the same time that it is persistently or predominantly higher than short-run marginal cost by a much larger percentage, then I must respectfully protest. The point involved is an important one and the subject of quite widespread confusion, so I should like to spell out its implications.

First, in an industry that is really purely competitive, there is no need to indulge in any apologetics about short-run equilibrium. In any purely competitive equilibrium, whether short-run or long-run, price is always equal to the respectively relevant marginal cost. Furthermore, in long-run equilibrium, price is equal to both long-run and short-run marginal cost (provided only that they are both determinate). This follows, of course, from the tangency or envelope relationship between long-run and short-run total cost. Then, according as output in the short run is either above or below the long-run equilibrium level, short-run marginal cost is also either above or below long-run marginal cost, respectively. Thus it does not follow that short-run marginal cost is either necessarily or presumptively below long-run marginal cost, despite a widespread belief to the contrary even among theorists who should know better. Accordingly, in any evaluation as to how "nearly" purely competitive a given industry may be, the relevant test is how nearly price corresponds to short-run marginal cost on the average through time. Therefore, an industry or firm which keeps price not too far above some version of long-run marginal cost on the average, but which consistently exhibits an appreciably higher excess of price over short-run marginal cost, should not get very good marks as to the closeness of its approximation to pure competition. Under these conditions, in other words, there is an unflinching indication that capacity is persistently in excess of what it should be for either purely competitive or efficient equilibrium. On the other hand, as already suggested above in connection with Hayek's observation about

cost inflation, this is by no means an unusual case under conditions that are other than purely competitive.

Stigler's estimate as to the number of industries that adequately approximate pure competition or, as he puts it, "in which monopoly power is inappreciable" (p. 47), is further biased by his willingness to certify an industry as "competitive" if either of two conditions are met. The first concerns an industry's structure and the second concerns its performance (p. 47): "(1) It must possess a large number of independent firms, none dominant in size, and additional firms can enter the industry. (2) It may possess few firms, or a few firms of relatively large size, but the departures from a competitive situation are not large." I have already indicated my misgivings as to the excessive leniency of the second condition, since it rests on the "approximate equality of price and marginal cost in the long run." Furthermore, the first condition is also too lenient, in that it allows an industry to qualify as "competitive" even if the gap between price and marginal cost is substantial. This may well be the case (1) because product differentiation alone is capable of raising price appreciably above marginal cost, even when the rival firms are numerous and entry is easy, and (2) because local oligopolistic relationships may also exist, even when firms are numerous throughout a broadly and loosely defined "industry."

Let me conclude with the almost obligatory reference to what remains for future research. Despite the methodological differences just indicated, it still seems to me that the basic conflicts between Cambridge and Chicago in this area are primarily factual, at least to the extent that we seem to agree that the relationships of price and marginal cost are of central importance. What we ideally need, accordingly, is a kind of profile of price, marginal-cost, and value-added relationships throughout the economy. Even if the whole job would be a staggering one, it would still be worth while to have it done if only on a piecemeal sampling basis. Only then would we really be in a position to evaluate definitively, for example, Stigler's conclusion that the U.S. economy is about 70 percent "competitive" or Harberger's ingenious but extraordinarily naïve estimate that resource misallocation in American manufacturing entails a cost of only about \$2.00 per capita.<sup>5</sup> My own hostage to fortune is that the relevant data would reveal the disparities between price and marginal cost throughout the U.S. economy to be far greater than they are thought to be in Chicago.

<sup>5</sup> A. C. Harberger, "Monopoly and Resource Allocation," *A.E.R.*, May, 1954, pp. 77-87, esp. p. 84.

# MONOPOLISTIC COMPETITION AND WELFARE ECONOMICS\*

By WILLIAM J. BAUMOL

*Princeton University*

Definitive reappraisals of a major theoretical contribution are like a soprano's final farewell appearance. They come frequently and sometimes with monotonous regularity. It would not be easy to estimate the number of conclusive evaluations of Professor Chamberlin's contribution [4] but their rate of appearance seems recently to have been stepped up. This paper will no doubt quietly take its place in this massive and spontaneous tribute to the impact of the *Theory of Monopolistic Competition*.

It must be emphasized here that even the writings of those who would minimize the contribution of monopolistic competition, by their very number and vociferousness combine to constitute an encomiast, as it were, praising by strong (and frequent) damns.

To deal with the subject which has been assigned to me—the contribution of monopolistic competition to welfare economics—I feel I must probe into ancient history. For in recent years it has become unfashionable for the welfare economist to speak about anything so substantive. Rather, he has grown exceedingly introspective and has been spending much of his time determining the circumstances under which he has any right to speak at all. No longer does he ask whether monopolistic competition is good for the country, but tries, instead, to construct criteria of increasing subtlety which, if the circumstances were right, and the appropriate data and value judgments were available, would permit him to break his stubborn silence on more applicable matters. Fortunately, monopolistic competition theory antedates this period and so we can be somewhat sanguine as we proceed to look at the literature.

## *I. Monopolistic Competition and the Allocation of Resources*

One of the main substantive areas of contribution of welfare theory is its analysis of the allocation of resources under different market structures and its comparison with the social optimum—the so-called “ideal output.” However, most results which have emerged have dealt with the case of universal pure competition or with the presence of a

\* The author would like to express his thanks to the National Science Foundation whose grant greatly facilitated the completion of this paper.

single monopolist in a world otherwise unsullied by the presence of monopolistic elements.<sup>1</sup>

One need not seek far for the reason monopolistic competition has failed to make a stronger showing here. Resource allocation among industries and among products is, after all, intrinsically a matter for general equilibrium (or general disequilibrium) analysis. One cannot say in absolute terms that the output of a single product, taken in isolation, is ideal. Only if its producers have kept out of the hands of the makers of other commodities neither too large nor too small a collection of inputs can we say that their output levels are consistent with maximal social well-being. In other words, to pass judgment on the allocation of resources one must be able to account for the interrelationships among the outputs of the various industries. But it is here that monopolistic competition is notoriously weak. Despite Professor Triffin's noble attempt, it must be recognized that general equilibrium theory remains, by and large, a theory of pure competition.

## II. *Product Differentiation and Resource Allocation*

The case of product differentiation has proved particularly resistant to general equilibrium analysis. Where we can propose no satisfactory criterion by which one can recognize either an industry or a product, it is perhaps too much to expect that we can say much by way of rigorous analysis about the allocation of resources among industries and products. The theory of product differentiation has been and remains a theory of the firm.

## III. *Oligopoly and Resource Allocation*

Rather more hopeful is the case of the oligopolistic industry whose product is homogeneous. And, indeed, the literature does contain models which offer some (rather limited) welfare implications. For example, the Cournot theory tells us that oligopolists who behave in accord with its model will produce an output smaller than that which would be yielded by a comparable competitive industry. From this we may infer that in suitably specified circumstances oligopoly will lead to a misallocation of resources and an inadequate output of the oligopolized product.

But when we recognize the number of conditions which are required by this argument we are left deriving small comfort from it. Not only must our industry be composed exclusively of unreconstructable Cournot oligopolists who never learn by experience that their output decisions affect those of their rivals, but all other industries must be

<sup>1</sup>Two noteworthy exceptions are Chap. 27 in Mrs. Robinson's classic volume [9] and R. A. Berry's unpublished dissertation [3].

uncontaminated either by elements of monopoly or by external economies and diseconomies. In Mrs. Robinson's "world of monopolies" all bets are off. For with a given level of employment of resources (the fundamental assumption of allocation theory) it is impossible for all industries simultaneously to use too small a proportion of society's resources. If all industries demands for resources are weaker than would be those of their competitive equivalents, the allocation of resources may nevertheless conceivably remain unaffected,<sup>2</sup> and in any event it is not easy to predict or evaluate such changes in allocation as will occur.

Moreover, in such a world our efforts at prescription are further hampered by the exacting requirements of the Lipsey-Lancaster theory of the second best. For they have argued that in circumstances in which some necessary conditions for an optimum are violated, there is little to be said for enforcement of the remaining necessary conditions. It is not difficult to construct a persuasive resource allocation illustration for this view. In a two-industry world, as we have seen, if both industries are monopolized, there is no reason to assume that a serious misallocation of resources will result. But enforcement of marginal cost pricing on one of these industries is a sure way to produce a maldistribution of resources.<sup>3</sup> It is at least arguable that the relatively competitive circumstances of American agriculture in pre-New Deal days aggravated significantly the misallocation of the nation's inputs.

#### IV. *Sales Maximization and Resource Allocation*

One somewhat personal postscript may perhaps be permissible before leaving our discussion of allocation. In recent months a number of prominent economists and I have received an unpublished paper in which my sales maximization hypothesis ([2], Chaps. 6-8) is denounced as an attack on the profit system. I must admit that this rather took me by surprise because I had been more fearful that my hypothesis would be considered by some to be a piece of capitalist apologetics. For it seems clear enough from the preceding discussion, and indeed it is a commonplace, that profit maximizing behavior in the presence of monopoly elements can frequently yield results which are not optimal from the point of view of society as a whole.

By contrast, if oligopolists seek to maximize their sales (total revenue), the consequences may be somewhat more commendable. This is suggested by a crude view of the matter, utilizing the theorem that sales (total revenue) maximization will yield outputs larger than those which maximize profits. For, if the trouble with profit maximizing

<sup>2</sup> For a more detailed argument see Berry [3], Chap. II.

<sup>3</sup> Cf. Berry, *loc. cit.*

monopolistic and oligopolistic outputs is that they are too small, then (if only they do not overcompensate) the larger sales maximizing output levels have some presumption in their favor.

A somewhat more subtle argument (which should not be taken very seriously but is perhaps a bit more amusing) takes as its view that sales maximizing oligopolists seek to return to their stockholders no more (or less) than "normal" profits; i.e., the rate of interest on capital plus some compensation for risk. If this were so and if all firms were to succeed in this goal, then a mixed world of competitors and oligopolists might yet produce the ideal output, provided, of course, that it were untouched by externalities.

#### *V. The Excess Capacity Theorem: Interpretation*

The analysis of product differentiation, while it has had little to say about the allocation of resources among industries, has yielded one celebrated theorem with much noted welfare implications. This is Professor Chamberlin's excess capacity theorem which rests on the argument that high profits will lead to the manufacture of new substitute products and that a sufficient rate of entry can force the total elimination of profits. Since this requires tangency between the company's average cost and revenue curves and since the demand (average revenue) curve will have a negative slope under monopolistic competition, the equilibrium point must lie on a negatively inclined portion of the average cost curve; i.e., the company must be able to reduce its unit costs by increasing its production.

Mr. Kaldor [7] has long ago discussed, in some detail, the plausibility of the assumptions which lie behind this argument, and in a similar spirit Professor Machlup ([8], pp. 311-44) has proposed several alternative interpretations of the theorem:

1. Tautological interpretation: The profits of the firm under monopolistic competition must be imputed to one or more of its inputs, and if this rent is included in cost, the company's net profit must necessarily be zero. The tangency result and the excess capacity theorem therefore follow inescapably. It is worth noting that Mrs. Robinson's version of the theorem is of this variety, which, as Machlup has shown, is not entirely without welfare implications, despite its tautological character.

2. A second variant of the theorem asserts that it is valid only where firms are sufficiently similar, small, and numerous. This interpretation restricts the range of relevance of the theorem very severely, not only because such circumstances are far from universal, but also because then the typical company's demand curve may plausibly be expected to be quite flat and excess capacity correspondingly insignificant.



3. A final interpretation of the theorem treats it as an approximation which indicates a tendency expected to be observed in most circumstances characterized by product differentiation. This is by far the most attractive variant of the theorem but it then becomes a matter for empirical verification.

Returning now to the tautological version of the theorem, we note that Professor Machlup has divided the rents imputed to the profit yielding inputs into those which are socially "allowable" and those which must be rejected because they represent no real significance for the social product. For example, if the firm enjoys a monopolistic advantage because it employs a really scarce input (such as a uniquely advantageous site), then this input service constitutes a very real advantage to the firm and to the community, and the rent is accordingly considered an allowable element in the relevant social unit cost curve. On the other hand, where the firm's product is differentiated by an ability to create an illusory quality advantage, e.g., by salesmanship, then no social rent can be imputed to the input in question and the social unit cost curve may fail to be tangent to the average revenue curve, so that equilibrium need not involve declining average social costs (there need be no excess capacity). Professor Machlup concludes rather despondently that we are then damned in the one case and damned in the other. In the one circumstance, where the rents are "real," excess capacity must also be very real; while in the other eventuality, where we may escape true excess capacity, we are likely to become tributaries to the artful producer of artificial monopolistic elements.

#### *VI. The Excess Capacity Theorem: Validity of Its Welfare Implications*

Recently, the very validity of the welfare implications usually drawn from the excess capacity theorem has come under attack, even in circumstances satisfying the assumptions of the theorem. It is appropriate to review and evaluate the objections which have been raised. Harold Demsetz [5] has argued that the theorem is undermined by the multiplicity of variables which enter the profit function under monopolistic competition. Professor Chamberlin's argument shows merely that a small increase in output from its equilibrium level must reduce average cost only if all other things remain constant. But, says Demsetz, other things cannot be expected to remain equal. Self-interest will usually induce managements to reconsider their advertising outlays whenever they add to their outputs. The net effect of a combined change in advertising and production levels cannot be predicted from the Chamberlinian analysis.

Suppose this much is granted. Demsetz on his side readily agrees that the *ceteris paribus* consequence of an output change from an equilibrium point will reduce unit costs. The question to be settled then, is, which assertion has the action—which of them is pertinent for welfare analysis?

Here, it seems to me, the logic of the matter rules against Demsetz. If any partial derivative of the average cost function is negative at the equilibrium point, we cannot be at an optimum. Lower unit operating costs are necessarily possible and the productive arrangements must necessarily be wasteful in the sense which will presently be specified. True, if the firm is left to adjust by itself to a decree which requires it to increase output, the result may be even worse than it was initially. If it increases its advertising along with its production, its unit costs may well rise. But this does not mean that the initial equilibrium, however much less costly it may have been, was sufficiently large to yield minimum costs. On the contrary, granting its premises, the Chamberlin argument shows conclusively that his equilibrium point is characterized by unused opportunities for cost reduction.

A second attack on the validity of the excess capacity theorem was mounted by G. C. Archibald ([1], pp. 18-19). Archibald also rests his case on the multivariate character of decision making under product differentiation. But he argues that the average cost curve which Chamberlin shows to be downward sloping at the equilibrium point is an aggregation of production and advertising outlays. However, this negative slope is perfectly compatible with rising unit output costs, provided only that advertising costs are falling with sufficient rapidity to offset any rise in per-unit production outlays. "Hence," he concludes, "when advertising is introduced, tangency is consistent with production at less than or more than minimum cost of production." One can hardly quarrel with the logic of Mr. Archibald's argument, and it must be emphasized that he carries it no further and makes no attempt to draw any welfare implication from his brief discussion.

One may, then, agree with him and yet simultaneously warn the reader against jumping to conclusions about the consequence of his analysis. For, so long as the combined curve of output and advertising costs slopes downward, it surely follows that an increase in the firm's production level can reduce the real resources tied up in a unit of its product, whether these resources are expended in the factory or by way of Madison Avenue. Chamberlin's negative slope remains a strike against the efficiency of unregulated product differentiation.

Very recently, Professor Friedman has also loosed one of his formidable thunderbolts upon the theorem. He argues ([6], p. 67):

In a world of specialized resources, and certainly in one of monopolies whatever their source, [average cost equals average revenue] is either a convention of accounting (total

costs equal total receipts) or, if defined to be an equilibrium condition, is imposed on the firm by the capital market via the revaluation of specialized resources. This is what renders the tangency condition irrelevant to productive adjustments and makes the alleged implications about "excess capacity and unexploited economies of scale"—with or without advertising—highly misleading if not downright wrong.

Professor Friedman, of course, has a valid point when he later expresses his dislike of the use of the term "capacity" for the minimum point on the average cost curve. But the fact remains that the excess capacity hypothesis (note I now do not call it a theorem) if not interpreted in accord with the Kaldor-Machlup tautological version, can still indicate that all is not quite for the best in a best of all possible economies. If small firms with heterogeneous products do, in fact, tend to drive one another's profits down to zero with no monopoly rents included (and as one who grew up among small retailers I can only consider this an understatement), then very likely each firm will be left so small a segment of the market that it must proceed at an uneconomically small scale of operation. Crude observation suggests that this phenomenon occurs all too often and that the tangency analysis can offer a meaningful hypothesis to explain it. The essence of the matter is an empirical hypothesis about the nature of entry under monopolistic competition, which, if valid, will prevent any substantial monopoly rents and hence any material revaluation of resources by the capital market. Of course, observation may show that the required entry behavior very rarely or perhaps, even, never occurs. But then the model will be shown to be false, but it still will neither be empty nor devoid of welfare implications.

### VII. *Standard Caveat re the Welfare Implications*

There is little need to review in detail the standard reservations regarding the welfare implications of the excess capacity theorem. We have often been warned of its pitfalls and the caution which is required in its interpretation. The two main points in this matter may quickly be summarized:

1. The excess capacity theorem is not a statement about the desirability of the allocation of resources among industries. It does not say that there will be too little produced by an industry (however defined) whose products are differentiated. Rather, the theorem tells us that the organization of the "industry" into firms is apt to be wasteful. It suggests that the same total output if produced by a smaller number of more sizable firms, can be provided at a lower real cost per unit, and hence a smaller total use of society's scarce resources.

2. The second relevant reservation refers to the social costs of standardization. If there is a reduction in the number of firms pro-

ducing differentiated outputs, the variety of products available to the consumer must fall. Whether the resulting saving in resources is then to be considered a net gain is a matter of judgment which will doubtless vary from case to case. In other words, the excess capacity theorem represents a real social cost only if the increased choice which it offers consumers does not provide adequate psychic compensation for the added costs which it imposes.

### VIII. *Measures for the Elimination of Excess Capacity*

Where excess capacity is considered undesirable, one may well ask what can be done about it. Aside from direct nationalization and abolition of excess capacity by fiat, an obvious means for going about it is a licensing system whereby the number of firms is restricted to any level deemed desirable by the authorities. Clearly such an enforced reduction in the number of companies can impart an upward shift to the demand curve of each surviving firm, and by skillful choice of the number of licenses issued the marginal revenue curve can be forced near the minimum point on the original average cost curve where the corresponding marginal cost curve is always to be found. Hence the profit maximizing output will coincide with the competitive output level for the firm and excess capacity will have been eliminated.

One new result in this area also merits some note. Mr. L. G. Sandberg [10] has recently shown in a neat argument that price ceilings can be used to induce monopolistic competitors to produce "competitive" (least average cost) outputs. It is merely necessary to set the price ceiling at the level of minimum average cost. At any other output, firms must, then, lose money. Suppose the demand before the typical firm is insufficient to permit it to sell the least average cost output. Some companies will then be forced to leave the field and the demand curves of the remaining firms will shift to the right. This process will continue until each remaining firm can just sell the least cost output and there the equilibrium point will be found.

### IX. *Concluding Comments*

This seems to be about all we can say about the specific welfare implications of the theory of monopolistic competition. It may well seem to the reader that more might have been expected especially in the first flush of enthusiasm for the newborn and more general model of the firm. But it must be remembered that the paucity of welfare results may be as much a commentary on the state of welfare theory as on the theory of that monopolistically competitive firm.

More important, it should be recognized that some of the most significant consequences of Chamberlin's contribution are less tangible

and come by way of its negative implications. One prominent member of our profession has described Chamberlin's book as the most influential single work ever produced by an American economist. The continuing impact of the book consists primarily in the habits of mind and point of view which it has engendered. Few economists today would accept, hastily and uncritically, welfare conclusions derived exclusively from a purely competitive construct. Moreover, we have learned to recognize a wide variety of types of monopolistic element and have at least formed intuitive judgments about their social consequences. No doubt, many of these judgments were implicit long ago in the thoughts and writings of some of the wisest members of our profession. But this detracts nothing from Professor Chamberlin's contribution. For it is the nature of a revolution in a body of analysis, that it renders tangible and readily available ideas which *dogmengeschichte* later shows to have been floating about well before the revolutionaries came into power.

## REFERENCES

1. G. C. Archibald, "Chamberlin versus Chicago," *Rev. of Econ. Studies*, Oct., 1961.
2. W. J. Baumol, *Business Behavior, Value and Growth* (Macmillan, 1959).
3. R. A. Berry, "Welfare Implications of Monopoly" (Ph.D. dissertation, Princeton, 1963, unpublished).
4. E. H. Chamberlin, *The Theory of Monopolistic Competition* (5th ed., Harvard Univ. Press, 1947).
5. Harold Demsetz, "The Nature of Equilibrium in Monopolistic Competition," *J.P.E.*, Feb., 1959.
6. Milton Friedman, "More on Archibald versus Chicago," *Rev. of Econ. Studies*, Feb., 1963.
7. Nicholas Kaldor, "Market Imperfection and Excess Capacity," *Economica*, N.S. Vol. II, 1935.
8. Fritz Machlup, *The Economics of Sellers' Competition* (Johns Hopkins, 1952).
9. Joan Robinson, *The Economics of Imperfect Competition* (Macmillan, London, 1933).
10. L. G. Sandberg, "A Note on the Effect of Price Controls on *The Theory of Monopolistic Competition*," *American Economist*, 1963.

## DISCUSSION

JESSE W. MARKHAM: Given the high perishability of books and ideas in the social sciences generally, the title of this program is itself a signal tribute to Chamberlin's *Theory of Monopolistic Competition*. This tribute is all the more impressive when we remind ourselves that the initial version of the work was the product of a graduate student, barely more than twenty-five years old, staking out in his thesis the relatively modest claim that his "hybrid theory affords a more illuminating approach to the study of the price system than does a theory of perfected competition, supplemented by a theory of monopoly" (5th ed., Preface, p. 41).

The authors of all three papers are in agreement that the tribute is richly deserved. Bishop ranks *Monopolistic Competition* in microeconomic theory with Keynes's *General Theory* in macroeconomics, attributes to it the demise of much economic illiteracy, including his own, and concludes that it has emerged virtually unscathed from frequent and heavy bombardment delivered from such close-range citadels as Cook County, Illinois. Bain considers it the parent that gave birth, with Edward Mason serving as midwife, to the field of industrial organization as we now know it; that is, a field of inquiry centered on Chamberlinian theory rather than institutional description. And Baumol, while concluding that the *Theory of Monopolistic Competition* has had relatively slight positive impact on general equilibrium theory—and hence on welfare theory—it stands as a formidable caveat to the uncritical acceptance of welfare conclusions derived from the constructions of pure competition.

The three papers correctly identify Chamberlin's models of the large-number case with tangency of average cost and average revenue and the small-number case with and without product differentiation as his important contributions, and those that have evoked most of the discussion and controversy. Baumol devotes over half of his paper to the first of these, and its corollary, the excess capacity theorem. I have no quarrel with placing the two models on equal theoretical footing, but as an empirical matter it appears to me that a disproportionate amount of time and effort has been devoted to the large-number excess capacity model at the expense of neglecting the strictly oligopolistic case and its implications for welfare theory, general theory, and, most importantly, public policy. This observation rests largely on the proposition that what often appears from the product market to be a good candidate for the average-cost average-revenue tangency case on closer examination of the firm structure of the industry involved nearly always turns out to be oligopoly—both in structure and in most behavioral characteristics.

Advertising and research and development are the two principal means whereby a firm differentiates its product—advertising through directly altering consumers' preferences and research and development through product change and the development of new products. The level of concentration for both of these activities is high. The twenty-five largest national televi-

sion advertisers typically account for 32 percent of total national television advertising and a staggering 60 percent of total network time sales. The twenty-five include five pharmaceutical companies, three automobile manufacturers, four cigarette producers, eight food, cereal and dairy companies, and two electrical equipment and appliance companies; Gillette, American Home Products, and Scott Paper complete the list. Typically the concentration ratio for the industries these firms inhabit is high, in most cases as high as 75 percent. Similarly, the fifty largest employers of research and development personnel (scientists, engineers, and supporting laboratory technicians) account for one-third of the total R and D personnel employed. Empirical studies attempting to correlate R and D expenditures with levels of concentration have been subject to considerable statistical error, but there is little doubt that they are highly correlated. The evidence is persuasive that most of the product-differentiating activity takes place in a market environment of oligopoly, where the firm typically has a wider range of choice with respect to prices and output than that single point at which its downward sloping demand schedule is tangent to its average unit cost curve. Put another way, Ivory Soap, Ivory Flakes, Camay, Lava, Duz, Tide, Cheer, Dreft, Oxydol, Dash, Joy, Comet, Cascade, Spic and Span, and Zest comprise an impressive list of differentiated soaps, detergents, and cleansers, but they are all produced by Procter and Gamble, who accounts for 54.5 percent of the national total; Procter, with Colgate-Palmolive and Lever Brothers, its two largest rivals, account for more than 80 percent.

None of this is meant to imply that all the discussion over the past thirty years concerning the large-number excess capacity case has been intellectually unrewarding, but it does suggest that it may not have been highly relevant. The situation may be as much a statistical rarity as the perfectly competitive firm.

This conclusion is not entirely devoid of logical support. Advertising and research and development activities, the means whereby products are differentiated, are costly and constitute formidable barriers to entry. As Nicholls has pointed out in his study of the cigarette industry, the annual outlays on advertising by any one of the largest three cigarette producers amounted to several times the cost of an efficient cigarette plant; the tangency case assumes reasonably free entry. And the late Professor Schumpeter made innovational activity both a prerequisite to and a result of monopoly profits—profits which do not exist under average cost-average revenue equality.

Baumol concludes that "the continuing impact of [*The Theory of Monopolistic Competition*] consists primarily in the habits of mind and point of view it has engendered." His observation is of special validity in certain areas of public policy. In the 1920's a prominent economist testified before the Federal Trade Commission that since under competition the prices of all sellers were expected to be identical, it could logically be concluded from a factual showing of identical prices that competition prevailed. It is not likely that any economist—at least none of those who have read Chamberlin's excellent discourse on what we now call "conjectural interdependence" and what the legal profession calls "conscious parallelism"—has been guilty of

asserting this confused identity within the past thirty years. This accomplishment alone should place Chamberlin's book high on the list of great books of this century.

PETER O. STEINER: If economists had a trade journal with the flair and flavor of *Variety*, tomorrow's edition might well carry the story of this session under the headline: *BBB grade EHC AAA after XXX*.

The verdict is one in which most of us concur, and the question is why. What was the anatomy of this revolution that changed our view of price theory, of the nature of the market process, of the ways to study markets, and the ways to evaluate market performance? Though the *Theory of Monopolistic Competition* was an ambitious book, bold in tone and revolutionary in intent, its own central models were primitive, subject to flaws and in many ways mere baby steps away from the competitive schema which had such a grip on the profession. Yet almost at once these steps found paths, and while our backs were turned most of them led to progress too rapid to fit the notion of mere giant steps. Why and how did it happen?

Bain suggests, correctly in my view, that the essential contribution was the induced generation of a variety of market structures among which behavior and performance would differ and which cried out for both theoretical elaboration and empirical examination. But how did it happen? This was, after all, a book in pure theory by a man who was then and is now neither an institutionalist nor an empiricist by instinct, by training, or by design.

Looking at the theory, Bishop finds the merits and positive significance essentially in the proposition that Chamberlin gave us a demand curve for the firm with negative slope. By so doing, according to Bishop, he made us able to hold up our theoretical heads in the face of the real life existence of such phenomena as wages not equal to the value of marginal product, the existence of decreasing cost, "salesmanship," sticky prices and so on. Further, by providing a happy rationalization for such things he provided a degree of contact with such forms of reality as retail trade, advertising and the like.

This view seems to me substantially wrong and an impediment to both understanding and appreciation. Chamberlin did not invent marginal revenue but he did much more. On the relevance of monopoly elements, Pierro Sraffa told us in 1926: "It is necessary to abandon the path of competition and turn in the opposite direction, namely toward monopoly."

Cournot, one hundred and twenty-five years ago, made a valiant—and so often misunderstood—attempt to provide a rationale for a monotonic decrease of price from the monopoly price level to the competitive one as the number of sellers increased from one to infinity.

Chamberlin resurrected Cournot and made us familiar with him and his critics. He then liberated us once and for all from the one-dimensional view of market structure that proves so confining. I find quite extraordinary Bishop's closing statement that "we seem to agree that the relationships of price and marginal cost are of central importance. What we ideally need, accordingly, is a kind of profile of price, marginal-cost, and value-added relationships throughout the economy." Given that profile, we should want to explain—



assuming we are scientists rather than dichotomists—the variance that surely exists by reference to a substantial list of explanatory variables. Chamberlin's basic contribution—in my view—lay in suggesting primitively (which is not to say naïvely) what these variables might be.

The pre-Chamberlinian distinction between the familiar polar cases of monopoly and competition rested basically upon the single dimension of the number of sellers. The first step—and it has proven a giant step—was merely to relabel this dimension as the degree of recognition of interdependence.<sup>1</sup> This might, of course, be associated with the number and size distribution of sellers, but other variables might affect it as well. Once it is clear that it is the perceptions of sellers, not their numbers, that are involved, it matters little that Chamberlin's own models dealt again with polar cases. (His  $DD'$  demand curve represented complete recognition; his  $dd'$  curve, complete non-recognition.) In retrospect, the distinction between his “mutual dependence unrecognized” case—monopolistic competition—and the competitive one seems smaller than we once thought, but the importance of the behavioral variable does not. Perhaps the first and most direct descendant was the kinky demand curve (one branch  $DD'$ , the other  $dd'$ ), but no less direct are Fellner's examination of the tendencies toward and away from the joint monopoly (complete recognition) position and the theoretical work on cartels, price leadership, and tacit collusion. Most recently, within the context of game theory, there has been even further exploration of a variety of kinds of responses, perceptions, and consequences.

Chamberlin's second step—and this certainly appeared a baby step at the time—was to recognize that the classical polar cases involved polar assumptions about entry: free and instantaneous in the one case, totally blockaded in the other. Chamberlin kept the dichotomy but added the possibility of interdependence recognized with free entry (the “high” tangency position) as distinguished from either polar case, in terms of the nature of the results. The condition of entry as a genuine variable, alternative policies that firms might adopt toward it, the significance of leads and lags in entry, and the consequences of asymmetries in entry and exit, have been among the most fruitful follow-ups in the years since 1933, particularly in the deft theoretical and empirical work of Joe Bain.

The third dimension is, of course, the nature of the product. I tend to the position of Chamberlin's critics that his theoretical use of this concept is overrated: the large-numbers case with free entry and no recognition of interdependence is only trivially different from pure competition. The only formal role that product variety plays is the slight negative slope it imparts to this perceived (but not realized) demand curve—and it is thus no different than Hotelling's distance, or transportation cost, etc. But after Chamberlin—and because of the extended discussion he gave these matters—one ignores the consequences of the kind and quality of product—and indeed the kinds of products of multiproduct producers—at one's peril. These are certainly es-

<sup>1</sup> “If sellers neglect both their indirect and their direct influence upon price, the outcome will be the purely competitive price, *regardless of numbers*.” (*Theory of Monopolistic Competition*, 5th ed., p. 54. Emphasis in the original.)

sential elements in understanding differences in market behavior and in evaluating market performance. Again we have progressed beyond the dichotomy between "undifferentiated" and "differentiated," but recognizing the attribute was a precondition of finding a variable.

Finally, Chamberlin explicitly, albeit extratheoretically, discussed the factor of uncertainty. Again he limited the discussion to "uncertainty on the part of one seller as to what the other is going to do," but he foreshadowed what I believe to be Fellner's great contribution to the theory of pricing: examination of the roles and consequences of uncertainty.

In short, I believe that the revolution engendered by the *Theory of Monopolistic Competition* lay not in the case of the same name, nor in the possibility of inequality at equilibrium between price and marginal cost, but rather in the insight: "Duopoly is not one problem, but several. The solution varies depending upon the conditions assumed."

We would reword it today. It is not duopoly but market structure, behavior, and performance that interest us. The "several" has become "many." The "solutions" turn out to be elusive and require a major blending of theoretical and empirical efforts. The conditions prevailing rather than the conditions assumed have become of major interest. But these changes are the outgrowths, not the negations, of the insight.

Bain has sketched, better than I could, the impact on the study of industrial markets. I would add only that the inseparable blending of a priori theory and data—the hallmark of economic theory today (the consequence of adding enough variables to deal with real problems)—reached price theory a decade before it reached elsewhere, and Chamberlin's revolution is responsible.

As to the welfare implications, I am almost as unhappy with the bulk of Baumol's paper as I am with Bishop's. He too emphasizes the small and misses the large. The real revolution again lies in the same multiplicity of market forms that made theory richer and measurement necessary. Given these, and given an appreciation of the theorem of the second best, some explicit definition of satisfactory market performance becomes the test not the automatic consequence of the form of competition. A priori welfare pronouncements become much less plausible and much less interesting—and this is as true of Chamberlin's "excess capacity theorem" as it is of the earlier notion that every departure from competitive equilibrium was a step away from that best of all possible worlds which was "more provocative of virtue than virtue herself."<sup>2</sup> In the dichotomous world of monopoly and competition, monopoly elements were either natural or insidious, and policy prescription was easy in either case. In the post-Chamberlinian world, life is harder. It is, I think, no mere coincidence that these same thirty years have marked the disappearance of the orators of elementary economics whose stock in trade was righteous indignation at the fools and scoundrels who frustrated our realization of an ideal world of competition. While this is one of the lesser contributions of a genuine revolution, for it alone we should be forever grateful.

<sup>2</sup> John Bascom, *Political Economy: Designed as a Text Book for Colleges* (1859).

# PRINCIPLES OF EFFICIENCY

## THE MEASUREMENT OF WASTE

By ARNOLD C. HARBERGER\*

*University of Chicago*

### I

The subject of this paper might be called "The Economics of the  $n$ th Best." This would distinguish the approach taken here from that taken by Lipsey and Lancaster in their fine article, "The General Theory of Second Best" [6], as well as from the conventional concern of economic analysis with the characteristics of fully optimal situations. To state the differences briefly, the conventional approach is concerned with how to get to a Pareto-optimal position, the Lipsey-Lancaster approach is concerned with how to make the best of a bad situation (i.e., how to get to a position which is optimal subject to one or more constraints which themselves violate the conditions of a full optimum), while this paper is concerned with measuring the deadweight loss associated with the economy's being in any given nonoptimal position.

The measurement of deadweight losses is not new to economics by any means. It goes back at least as far as Dupuit; and more recently Hotelling [4], Hicks [3], Debreu [2], Meade [10], and H. Johnson [5] [6] have made important contributions. Nonetheless I feel that the profession as a whole has not given to the area the attention that I think it deserves. We do not live on the Pareto frontier, and we are not going to do so in the future. Yet policy decisions are constantly being made which can move us either toward or away from that frontier. What could be more relevant to a choice between policy A and policy B than a statement that policy A will move us toward the Pareto frontier in such a way as to gain for the economy as a whole, say, approximately \$200 million per year, while policy B will produce a gain of, say, about \$30 million per year? What could be more useful to us as a guide to priorities in tax reform than the knowledge that the deadweight losses stemming from the tax loopholes (percentage depletion and capital gains) open to

\* I wish to express my indebtedness to Martin J. Bailey, Donald V. T. Bear, Milton Friedman, John Hause, and Harry G. Johnson for helpful discussions during the preparation of this paper. In addition, I am grateful to the members of the Workshop in Mathematical Economics and Econometrics at the University of Chicago for comments made at a seminar on this subject. Needless to say, all errors that remain are my own responsibility.

explorers for oil and gas are probably greater in total magnitude than the deadweight losses associated with all the other inefficiencies induced by the corporation income tax? What could be more tantalizing than the possibility (which I believe to be a real one) that the U.S. tariff, whose indirect effect is to restrict the equilibrium value of U.S. exports, produces by this route a gain for the U.S. from a partial exploitation of U.S. monopoly power in world markets which nearly offsets (or perhaps fully or more than fully offsets) the efficiency-losses produced by tariff-induced substitution of more expensive domestic products for cheaper imports? These and similar questions seem to me so interesting, so relevant, so central to our understanding of the economy we live in, that I find it hard to explain why the measurement of deadweight losses should be the province of only a handful of economists rather than at least the occasional hobby of a much larger group. Let me simply suggest four possible reasons for the apparent unpopularity of the loss-measurement game:

1. Even the simplest attempts to measure the deadweight loss (or, as I prefer to call it, the welfare cost) associated with particular distortions involve the use of numerical values for certain key parameters (elasticities of demand, of substitution, etc.), which may be impossible to obtain at all, or which may be estimated but with substantial error. Workers in this field must be ready to content themselves with results that may be wrong by a factor of 2 or 3 in many cases. But, on the other hand, it is a field in which our professional judgment is so poorly developed that the pinning down of an answer to within a factor of 2 can be very helpful. Be that as it may, one cannot expect the field to attract colleagues who prefer their results to be meticulously exact.

2. While it is relatively easy to measure the welfare costs of a particular distortion when one assumes other distortions to be absent, it is much more difficult to carry through the measurement in a way which takes account of the presence of other distortions. One of the profound lessons taught us by earlier workers in this field (Hotelling [4], Viner [11], Lipsey-Lancaster [7], Corlett and Hague [1], Little [8], and others) is that an action (i.e., imposing a tax of  $T_1$  per unit on good  $X_1$ ) which would take us away from a Pareto optimum if we were starting from that position can actually bring us toward such an optimum if we start from an initially distorted situation. Crude measures can thus mislead us, while correct measures are hard to come by.

3. Many people find it difficult to isolate the measurement of efficiency-losses due to particular distortions from the changes in the distribution of income that they conceive would ensue if the distortions were actually removed. Of these, some are undoubtedly not willing to make the kind of assumptions they have to make in order to compare the changes in welfare of different individuals or groups.

4. Consumer surplus, in spite of its successive rehabilitations, is still looked upon with suspicion by many economists. In spite of the fact that it is possible to formulate measures of welfare cost which do not directly involve the use of the consumer surplus concept, the most convenient and most frequently cited measures of welfare cost do involve this concept. Thus, I venture to guess, another group of potential workers (or at least tasters) in the vineyard do not venture to enter.

## II

The main purpose of this paper is to explore a variety of possible ways of formulating measures of deadweight losses. All the ways considered are members of a single family. This section begins by expounding a widely accepted approach to the problem and then proceeds to extend this approach to what I believe are new areas.

Let us begin by assuming that the only distortions present in the economy are taxes. Monopoly elements, externalities, and other market imperfections will be introduced at a later stage. We shall assume that the economy will seek and find a unique full employment equilibrium once its basic resource endowments, the distribution of income, the quantities of goods purchased by the government, and the set of distortions (taxes) are known. Letting  $X$  represent the vector of equilibrium quantities,  $D$  be a vector representing the proportion of total income received by each spending unit,  $G$  be a vector representing the quantities of the different goods and services purchased by government, and  $T$  be a vector representing the tax levied per unit of the different goods and services produced in the economy, we have  $X=f(D, G, T)$ .

Now to isolate the efficiency effects of distortions, we must hold  $D$  and  $G$  constant. Thus, with respect to  $D$ , we conceive of the possibility of keeping the percentage share of each spending unit in the total national income constant by means of neutral taxes and transfers. With respect to  $G$ , we assume that, in any pair of situations being compared, the government buys the same bundle of goods and services. Even though the comparison of two actual situations might be between  $X=f(D, G, T)$  and  $X'=f(D', G', T')$ , we split up the move from  $X$  to  $X'$  into a minimum of two steps. The first step is from  $X=f(D, G, T)$  to  $X^*=f(D, G, T')$ . This step isolates the efficiency aspects of the change. The move from  $X^*$  to  $X'$  entails no change in the distortions affecting the economy, and involves only shifts in the distribution of income and in the level of government expenditures. To the extent that the tax yield produced by the vector  $T$  is insufficient or more than sufficient to finance the expenditure vector  $G$ , we assume that neutral taxes or transfers will be called upon to make up the difference. (Should fiscal policy

measures be necessary to provide full employment, neutral taxes and transfers would be the instruments used to bring the total tax take to the required level. Government expenditures, on our assumptions, would be held fixed.)

The above assumptions have the effect of setting first-order income effects (whether caused by redistributions or changes in the size of government purchases) to one side so as to isolate the efficiency effects of alternative tax patterns. They put us in a world of substitution effects and of relative prices. When dealing with relative price phenomena, it is customary to treat a single product as the numeraire. This procedure is, however, not essential. One could normalize by holding any desired index of prices constant, or in a variety of different ways. For our purposes, it is convenient to normalize by holding the money national income constant as among all possible situations being compared. We could alternatively hold constant money net national product, gross national product, gross national product less excise taxes, or any of a variety of other possible aggregates. But, as will be seen, holding money national income constant is exceedingly convenient for the problems with which we shall deal.

Let us consider first a case that has been frequently dealt with in the literature. Assume that the production function of the economy is linear and that only one factor of production, in fixed supply, is involved in production.

The fact that the only distortions present in our system are per unit excise taxes assures us that when the vector  $T=0$ , we are at a Pareto optimum. (In this case the government is raising all its revenue by taxes that are by definition neutral; e.g., head taxes.) Thus if we set up an index of welfare  $W$  as a function of the tax vector  $T$ , we have that  $W_{\max}=W(0)$ . We can take money national income,  $\bar{Y}$ , as the measure of  $W(0)$ . We can, therefore, indicate the level of welfare associated with any tax vector  $T$  by  $\bar{Y}$  plus a deviation  $\Delta W$ , depending on  $T$  and expressed in the same units as  $Y$ . The relevant expression for  $\Delta W$ , in a wide class of cases, is

$$(1) \quad \Delta W = \sum_{i=1}^n \int_0^{T_i} \sum_{j \leq i} T_j \frac{\partial X_j}{\partial T_i} dT_i,$$

Two basic rules underlie this expression.

First, if as a result of an increment  $dT_i$  in the unit tax on  $X_i$ , there is an increment or decrement  $dX_j$  in the equilibrium quantity of a good  $X_j$  in the market for which no distortion exists, the change  $dX_j$  carries with it no direct contribution to the measure of  $\Delta W$ . For each successive minute increment of  $X_j$ , demand price is equal to marginal cost,

and the gain to demanders of having more of  $X_i$  is just offset by the costs of producing the extra amount.

Second, if as a result of an increment  $dT_i$  in the unit tax on  $X_i$  there is an increment  $dX_j$  in the equilibrium quantity of good  $X_j$ , in the market for which a distortion  $T_j$  already exists, there is a social gain associated with the change  $dX_j$  equal to  $T_j dX_j$ . Here demand price exceeds marginal cost, on each unit increment of  $X_j$ , by the amount  $T_j$ . Likewise if  $dX_j$  is negative, there is a social loss involved equal in magnitude to  $T_j dX_j$ .<sup>1</sup>

Obviously, the second rule given above contains the first, but I have set them out as two rules to emphasize the neutrality of changes taking place in undistorted sectors. Once this fact is appreciated, the rest of the road is easy.

Let me emphasize at this point that up to now there is nothing new in what has been said. Expression (1), and the rules behind it, say only that

$$(2) \quad \frac{\partial W}{\partial T_i} = \sum_j T_j \frac{\partial X_j}{\partial T_i}$$

This expression pops up in one form or another all through the literature on the measurement of welfare costs, the economics of second best, the theory of customs unions, etc. It appears, or can be derived from what appears, in Corlett and Hague [1], Hotelling [4], H. Johnson [5] [6], Meade [10], and Lipsey and Lancaster [7], among others.

Let us now linearize expression (1) by setting

$$\frac{\partial X_j}{\partial T_i} = R_{ji}.$$

With this substitution, (1) evaluates at

$$(3) \quad \Delta W = \frac{1}{2} \sum_{i=1}^n R_{ii} T_i^2 + \sum_i \sum_{j < i} R_{ji} T_j T_i$$

Expression (3) can be simplified, however, using the integrability condition

$$\frac{\partial X_i}{\partial T_j} = \frac{\partial X_j}{\partial T_i},$$

<sup>1</sup> Another way of looking at this problem is to consider that consumers, in transferring their demand to  $X_i$  are indifferent between what they get and what they give up for each marginal unit of purchasing power transferred; and that suppliers of factor services are likewise, for each marginal unit of services transferred, on the borderline of indifference. But if  $X_i$  goes up by  $dX_i$ , the government will obtain an increase in tax receipts of  $T_i dX_i$ , which (under our assumptions) will permit either a corresponding reduction in associated lump-sum taxes or a corresponding increase in lump-sum transfers. In short, "the people" gain to the tune of  $T_i dX_i$ .

which translates in the linearized form into  $R_{ij}=R_{ji}$ . In economic terms, this same condition derives from the fact that the welfare cost of a set of taxes should not, in a comparative static framework such as this, depend on the order in which those taxes are conceived to be imposed. Thus if we impose  $T_1$  first and follow it by  $T_2$ , we have  $\Delta W = \frac{1}{2}R_{11}T_1^2 + \frac{1}{2}R_{22}T_2^2 + R_{12}T_1T_2$ . If on the other hand we impose  $T_2$  first and follow it by  $T_1$ , we have  $\Delta W = \frac{1}{2}R_{22}T_2^2 + \frac{1}{2}R_{11}T_1^2 + R_{21}T_2T_1$ . Hence if the linearized expression (3) is to be invariant with respect to order of imposition of taxes,  $R_{12}$  must equal  $R_{21}$ , and in general  $R_{ij}$  must equal  $R_{ji}$ . This enables (3) to be simplified to

$$(4) \quad \Delta W = \frac{1}{2} \sum_i \sum_j R_{ij}T_iT_j.$$

For each  $R_{ji}(j < i)$  appearing in (3), we simply substitute  $\frac{1}{2}R_{ji} + \frac{1}{2}R_{ij}$ , to obtain (4).

A further condition on the  $R_{ij}$  can be established by noting that a set of taxes with some  $T_i \neq 0$  can at best produce an equal level of welfare as an undistorted situation. This yields

$$(5) \quad \Delta W = \frac{1}{2} \sum_i \sum_j R_{ij}T_iT_j \leq 0 \quad \text{for all possible values of } T_i, T_j.$$

As a special case of (5) we have

$$(6) \quad R_{ii} \leq 0 \quad \text{for all } i.$$

This is obtained when  $T_i \neq 0$ , while  $T_j = 0$  for all  $j \neq i$ .

We are by now quite close to establishing the Hicksian substitution conditions by the back door, so to speak. What we need to finish the job is the adding-up property. Suppose it to be true that a proportional tax at the rate  $t$  on all the  $X_i$  would indeed be a neutral tax. We can define  $T_i = t_i c_i$ , where  $c_i$  = marginal cost, and  $t_i$  = percentage rate of tax on  $X_i$ , to obtain:

$$(7) \quad \Delta W = \frac{1}{2} \sum_i \sum_j c_i c_j R_{ij} t_i t_j.$$

If an equal percentage tax on all commodities is neutral, we have

$$(8) \quad \sum_i \sum_j c_i c_j R_{ij} = 0.$$

But we actually have much more than this. If a proportional tax at the rate  $t$  is truly neutral, then, given our assumptions about the constancy of income distribution and of government purchases, it simply substitutes for the head tax that would have to exist if all the  $T_i$  were zero. It



must produce the same equilibrium quantity for each and every commodity. Thus we have that

$$(10) \quad \frac{\partial X_i}{\partial t} = \sum_j c_j R_{ij} = 0 \quad \text{for all } i.$$

This is the counterpart of the Hicksian adding-up property.

However, a tax at the rate  $t$  on all  $X_i$  will be neutral only in certain cases.

Case A) Suppose that, as was assumed above, the production frontier of the economy is linear—

$$\sum_i c_i X_i = \text{a constant.}$$

This means that total production is in inelastic supply, and therefore that a tax which strikes the value of all production at a constant rate will be neutral. In this case all the  $X_i$  must be final products; the  $R_{ij}$  here turn out to be precisely the Hicksian substitution terms.

Case B) Suppose that all the  $X_i$  are final products, and that the production frontier of the economy is convex from above. Suppose, moreover, that all basic factors of production are fixed in total supply. So long as a tax at the rate  $t$  on final products is in effect a tax at a fixed rate on the net earnings of all factors of production, it will be neutral, and condition (9) will hold. In this case, the  $R_{ij}$ , while obeying the properties of the Hicksian substitution terms, are actually quite different from them. Here the  $R_{ij}$  are really the "reduced form" coefficients showing how the equilibrium value of  $X_i$  (with supply and demand equal for all commodities) depends on  $T_j$ .

Case B presents no problem when capital is not among the basic factors of production, or when the relation between gross and net earnings of capital is the same in all uses. However, when capital is among the basic factors and when the relationship between gross and net earnings does (because of different depreciation patterns) differ among uses, then an equal tax on all final products will not be neutral, even though capital and other factors of production are fixed in total supply. This is because increases in the rate of proportional tax,  $t$ , will create incentives which would relatively favor the longer-lived applications of capital. An equal tax on value added in all industries, however, would be neutral in these circumstances, because we assume the net rate of return on capital to be equalized among all uses of capital. (This, of course, assumes that the stock of capital and the supplies of other basic factors of production are fixed.)

Problems quite similar to those presented by different depreciation

patterns in different applications of capital arise when the possibility of taxing intermediate products is introduced. As McKenzie [9] has forcefully pointed out, an equal percentage tax on all products will generally be nonneutral if any of the products in question are intermediate or primary products not in fixed supply. A tax at an equal percentage rate on value added in every activity will, on the other hand, be neutral so long as the basic factors of production are in fixed supply.

Case C) When considering taxes on value added we let  $X_i$  represent the volume of final product of industry (or activity)  $i$ ,  $v_i$  represent value added per unit of the product of activity  $i$ , and  $T_i$  represent the tax per unit of final product in industry. (Although the tax is levied on value added,  $T_i$  is here expressed per unit of product.) Once again letting  $R_{ij} = \partial X_i / \partial T_j$ , we have (5) as the expression for  $\Delta W$ . To reflect the neutrality of an equal percentage tax on value added everywhere, we require that the response of any  $X_i$  to such a tax be zero; i.e., that

$$\sum_j v_j R_{ij} t_j = 0 \quad \text{when} \quad t_j = t \quad \text{for all } j.$$

Here  $v_j$  = value added per unit of the product  $X_j$ ,  $t_j$  = percentage rate of tax on value added in industry  $j$ ,  $v_j t_j = T_j$ . Hence we have

$$\sum_j v_j R_{ij} = 0; \quad R_{ij} = R_{ji}; \quad \sum_i \sum_j R_{ij} T_i T_j \leq 0 \quad \text{for all } T_i, T_j; \quad R_{ii} \leq 0 \quad \text{for all } i$$

as before.

Case C deals rather neatly with problems of differential depreciation and taxes on nonfinal products. However, case C assumes that indirect taxes are levied on value added, whereas most frequently in the real world they are levied on the final product.

Fortunately, it is possible to translate product taxes into value-added taxes, and still stay within the framework of case C so long as inputs other than labor and capital enter their respective products in fixed proportions. The reason for this is obvious. All the effects of a tax at the rate  $t_i$  on product  $i$  can be replicated by a tax at the same rate on all factor shares (including materials input) entering into the production of product  $i$ . These are simply two ways of imposing the same tax. Suppose that with a tax of 10 percent on all factor shares in the  $i$ th industry an equilibrium is reached in which materials inputs account for half the value of product and labor and capital the other half. So long as materials inputs must be used in fixed proportions per unit of product, a shift from a 10 percent tax on all factor shares in the  $i$ th industry to a 20 percent tax on value added in the  $i$ th industry would introduce no incentive to change the equilibrium reached with a 10 percent tax on all factor shares. Purchasers could pay the same price for the product;

labor, capital, and materials sellers could get the same net reward; and the government could get the same tax take. Moreover, since the taxes on labor and capital shares would still be at equal percentage rates, there would be no inducement for substitution between them. In short, so long as materials are used in fixed proportions to output, we can translate any given tax on output into a tax on value added that is equivalent in all respects relevant for this analysis.

We now turn to a broader set of problems—all of which take into account the possibility of different rates of tax on the return to capital and to labor in any given activity. Consider first the set of possible taxes  $B_i$  per unit of capital in activity  $i$ . The change in welfare associated with such taxes can be written, assuming no other nonneutral taxes in the system, as

$$(11) \quad \Delta W = \frac{1}{2} \sum_i \sum_j G_{ij} B_i B_j,$$

where

$$G_{ij} = \frac{\partial K_i}{\partial B_j},$$

and  $K_i$  represents the number of units of capital employed in activity  $i$ . Correspondingly, if we consider the set of possible taxes  $E_i$  per unit of labor in activity  $i$ , and assume no other nonneutral taxes, we can write:

$$(12) \quad \Delta W = \frac{1}{2} \sum_i \sum_j M_{ij} E_i E_j,$$

where

$$M_{ij} = \frac{\partial L_i}{\partial E_j},$$

and  $L_i$  represents the number of units of labor employed in activity  $i$ . The  $G_{ij}$  and the  $M_{ij}$  will obey the following properties:

$$\begin{aligned} G_{ij} &= G_{ji}; & \sum_i \sum_j G_{ij} B_i B_j &\leq 0 \text{ for all } B_i, B_j; & G_{ii} &\leq 0 \text{ for all } i \\ M_{ij} &= M_{ji}; & \sum_i \sum_j M_{ij} E_i E_j &\leq 0 \text{ for all } E_i, E_j; & M_{ii} &\leq 0 \text{ for all } i. \end{aligned}$$

<sup>1</sup> We could here have explicitly set out an equation corresponding to (1); i.e.,

$$\Delta W = \sum_{i=1}^n \int_0^{B_i} \sum_{j \leq i} B_j \frac{\partial X_j}{\partial B_i} dB_i,$$

linearized this expression as in (3), and then used the symmetry property to obtain (11) or its counterpart. These steps are not presented explicitly in this and the other cases treated in this section.

Moreover, with fixed supplies of capital and labor we have

$$\sum_i G_{ij} = 0, \quad \sum_i M_{ij} = 0.$$

When nonneutral taxes are levied only on capital in different activities, (11) measures the cost of the distortions involved; when only labor is affected by nonneutral taxes, (12) is the relevant measure. But when nonneutral taxes are levied on both labor and capital in different activities, the interaction between them must be taken into account. Let us define

$$H_{ij} = \frac{\partial K_i}{\partial E_j} \quad \text{and} \quad N_{ij} = \frac{\partial N_i}{\partial B_j}.$$

Here symmetry exists between  $H_{ij}$  and  $N_{ji}$ . Suppose for example, we impose first a tax of  $B_1$  and then one of  $E_2$ . We obtain  $\frac{1}{2}G_{11}B_1^2 + M_{22}E_2^2 + H_{12}B_1E_2$  as our measure of  $\Delta W$ . If we conceive of  $E_2$  being imposed first, and then  $B_1$ , we obtain  $\frac{1}{2}M_{22}E_2^2 + \frac{1}{2}G_{11}B_1^2 + N_{21}E_2B_1$ . If we think of imposing a set of taxes  $B_i$  first and then a set of taxes  $E_i$ , we have

$$(13) \quad \Delta W = \frac{1}{2} \sum_i \sum_j G_{ij} B_i B_j + \frac{1}{2} \sum_i \sum_j M_{ij} E_i E_j + \sum_i \sum_j H_{ij} B_i E_j.$$

If we think of it the other way around, we have

$$(14) \quad \Delta W = \frac{1}{2} \sum_i \sum_j M_{ij} E_i E_j + \frac{1}{2} \sum_i \sum_j G_{ij} B_i B_j + \sum_i \sum_j N_{ji} E_j B_i.$$

For a reason that will be apparent later, it is most convenient to write:

$$(15) \quad \begin{aligned} \Delta W = & \frac{1}{2} \sum_j \sum_i M_{ji} E_j E_i + \frac{1}{2} \sum_j \sum_i N_{ji} E_j B_i + \frac{1}{2} \sum_i \sum_j G_{ij} B_i B_j \\ & + \frac{1}{2} \sum_i \sum_j H_{ij} B_i E_j. \end{aligned}$$

Now, when labor is in fixed supply, a tax on capital in industry  $i$  can only redistribute the existing amount of labor. Hence

$$\sum_j N_{ji} = 0.$$

Likewise, when capital is in fixed supply, a tax on labor in industry  $j$  can only redistribute the available capital, so that

$$\sum_i H_{ij} = 0.$$

The interaction terms disappear for neutral taxes because in this case  $E_i = \bar{E}$  for all  $i$ ,  $B_i = \bar{B}$  for all  $i$ . (Since the wage is assumed to be equalized in all uses of labor and since the net rate of return is assumed to be equalized in all uses of capital, an equal tax per unit of labor is also an equal percentage tax on value added by labor in different activities, and likewise for capital.)

Thus we have:

Case D) When labor and capital are in fixed supply, expression (15) measures the change in welfare due to any pattern of taxes on labor and capital in different activities. The  $M_{ji}$  and the  $G_{ij}$  obey the Hicksian conditions, with the adding-up property in this case

$$\sum_j M_{ij} = 0 = \sum_i G_{ij}.$$

All terms vanish for taxes on labor that are equal all uses together with taxes on capital that are equal in all uses. The interaction terms can in general be positive or negative, but the whole expression (15) must always be  $\leq 0$ . In this case the coefficients reflect not only conditions of final demand and supply but also conditions of factor substitution.

We now attempt to allow for the fact that the supply of labor in the market may itself be a function of the pattern of taxation. This question has been dealt with in the literature of second-best by Little, Corlett, and Hague, Lipsey and Lancaster, and Meade, among others. The key to at least the last three of these treatments is the substitution of the assumption (1) that the number of hours in the year is fixed for the assumption that the number of man-hours offered in the market is fixed. We can do this simply by adding another activity for labor—labeled “leisure” or “nonmarket activity.” If there are  $n$  market activities, we add an  $n+1$ st, and have

$$\sum_{j=1}^{n+1} L_j = \bar{L}.$$

This does not change the form of equation (12) but it does alter the definition of a neutral tax. Now an equal tax on all labor in market activities is not neutral, because it neglects the  $n+1$ st activity. However, a tax that struck all hours equally (including leisure hours) would be neutral. Hence we have

$$\sum_{j=1}^{n+1} M_{ji} = 0 = \sum_{j=1}^{n+1} M_{ij}$$

To measure the welfare cost of an equal tax of  $\bar{E}$  on all activities except leisure we take

$$(16) \quad \Delta W = \frac{1}{2} \sum_{j=1}^n \sum_{i=1}^n M_{ji} \bar{E}^2$$

but

$$\sum_{i=1}^{n+1} M_{ji} = 0, \quad \text{so} \quad \sum_{i=1}^n M_{ji} = -M_{j,n+1}.$$

Hence (16) reduces to

$$(17) \quad \Delta W = -\frac{1}{2} \sum_{j=1}^n M_{j,n+1} \bar{E}^2$$

But

$$\sum_{j=1}^{n+1} M_{j,n+1} = 0, \quad \text{so that} \quad \sum_{j=1}^n M_{j,n+1} = -M_{n+1,n+1},$$

Thus (17) reduces to

$$(18) \quad \Delta W = \frac{1}{2} M_{n+1,n+1} \bar{E}^2,$$

where  $M_{n+1,n+1}$  represents the responsiveness of leisure to a change in the tax rate on leisure (or to the negative of a change in the tax rate on work). This exercise illustrates, I think, the usefulness of properties of the kind that we have been establishing in the various cases examined. (16) taken by itself looks hard to interpret; with the aid of the adding-up properties, however, it can be reduced to (18), which is easy to interpret and perhaps even to measure.

The general expression for  $\Delta W$ , for a fixed capital stock and for a fixed amount of labor-plus-leisure, is

$$(19) \quad \Delta W = \frac{1}{2} \sum_{j=1}^{n+1} \sum_{i=1}^{n+1} M_{ji} E_j E_i + \frac{1}{2} \sum_{j=1}^{n+1} \sum_{i=1}^n N_{ji} E_j B_i + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n G_{ij} B_i B_j \\ + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^{n+1} H_{ij} B_i E_j$$

Its properties are basically the same as those of (15), modified only to take account of the fact that labor has  $n+1$  activities available to it while capital has only  $n$ . Thus, in the interaction terms we have

$$\sum_{j=1}^{n+1} N_{ji} = 0; \quad \sum_{i=1}^n H_{ij} = 0.$$

Hence we have:

Case E) Where capital is in fixed supply to market activities, but labor is in fixed supply only to market-plus-nonmarket activities, and

where taxes are considered which strike labor and capital differentially in different activities, (19) measures the change in welfare stemming from any set of such taxes. Neutral taxes in this case are taxes striking each unit of capital (or each dollar of net return from capital) equally, and taxes striking each hour of a worker's day equally. This last set of taxes could equivalently be called head taxes, but, as was shown above, convenient results can be obtained using properties derived from the neutrality of an equal tax per hour.

The formulation of case E is quite versatile. It can deal with proportional income taxation (equal percentage taxes on the income from labor and capital), and can recognize the nonneutrality of ordinary income taxation as regards the choice between labor and leisure. It can also cope with progressive income taxation, simply by using the effective marginal rate of tax to apply to income from labor and capital (here one has to assume that each individual's supply of capital is constant). It can cope with indirect taxes on intermediate as well as final products, provided that one is prepared to make the assumption that materials inputs bear fixed relationships to final products. And, most important of all, it can cope with property and corporation income taxes, which have widely differing burdens on the income from capital in different industries. Finally, it is possible to deal with situations in which all the above-mentioned taxes are simultaneously present, amalgamating those taxes (including allocations of excise tax receipts) falling on income from capital in each activity, and those falling on income from labor.

### III

This section consists of three "appended notes" to the earlier analysis. The first (A) reduces the expressions derived in Section II to a common simplified form. The second (B) discusses how distortions other than taxes can be incorporated in the analysis. The third (C) discusses the problems that arise when one eliminates the assumption of a constant capital stock.

A. The cases dealt with in the preceding section all have in common a simple property. Since, in (4)

$$\sum_j R_{ij}T_j \text{ can be expressed as } \Delta X_i,$$

(4) itself can be rewritten:

$$(4') \quad \Delta W = \frac{1}{2} \sum_i T_i \Delta X_i.$$

For cases A and B,  $T_i$  refers to taxes on final products only, and the  $X_i$ 's are final products. For case C, (4') might better be written

$$(4'') \quad \Delta W = \frac{1}{2} \sum_i v_i t_i \Delta X_i,$$

where the  $X$ 's are now final or intermediate products, the  $v_i$  represent value added per unit of product in activity  $i$ , and the  $t_i$  are percentage taxes on value added in activity  $i$ . Since

$$\sum_j G_{ij} B_j \text{ can be expressed as } \Delta K_i,$$

(11) can be written:

$$(11') \quad \Delta W = \frac{1}{2} \sum_i B_i \Delta K_i$$

Similarly, (12) can be written:

$$(12') \quad \Delta W = \frac{1}{2} \sum_i E_i \Delta L_i.$$

In (15),

$$\sum_i M_{ji} E_i + \sum_i N_{ji} B_i$$

can be expressed as  $\Delta L_j$ , while

$$\sum_j G_{ij} B_j + \sum_j H_{ij} E_j$$

can be expressed as  $\Delta K_i$ , so that (15) can be written:

$$(15') \quad \Delta W = \frac{1}{2} \sum_j E_j \Delta L_j + \frac{1}{2} \sum_i B_i \Delta K_i.$$

(15') also serves as an alternative form for (19), with the index  $j$  going from 1 to  $n+1$  and the index  $i$  going from 1 to  $n$ . Thus all of the cases discussed here are extensions of the "triangle-under-the-demand-curve" that emerges in textbook discussions of the excess burden of taxation. But I believe that for actual work the simplified forms presented above are not as useful as those presented in Section II of this paper, in which explicit account is taken of how the reaction coefficients  $R_{ij}$ ,  $G_{ij}$ ,  $H_{ij}$ ,  $M_{ij}$  and  $N_{ij}$  enter into the determination of the result. One can conceive, at least hypothetically, of measuring these reaction coefficients by experimental movements in individual tax rates. Once measured, they will enable us to estimate the changes in welfare associated with any arbitrary combination of taxes.



In practice one cannot expect to measure all the relevant reaction coefficients, but one can place reasonable bounds on their orders of magnitude and thus get estimates of the order of magnitude of the welfare costs of a given set of taxes, or of particular changes in the existing tax structure. In dealing with practical problems, the presumptive dominance of the diagonal elements in the matrices of reaction coefficients can be put to good use. Consider, for example, the case of a tax of  $T_1$  on  $X_1$ , in case A or B of Section II. If there are no other taxes present in the system, the change in welfare associated with this tax will be  $\Delta W = \frac{1}{2} R_{11} T_1^2$ . If there are other taxes already present in the system, the effect on welfare of adding a tax of  $T_1$  on  $X_1$  will be

$$(20) \quad \frac{\partial W}{\partial T_1} T_1 = \frac{1}{2} R_{11} T_1^2 + \sum_{i=2}^n R_{i1} T_i T_1, \text{ or}$$

$$(20') \quad \frac{\partial W}{\partial T_1} T_1 = \frac{1}{2} c_1^2 R_{11} t_1^2 + \sum_{i=2}^n c_i c_1 R_{i1} t_i t_1.$$

Since

$$\sum_{i=2}^n c_i R_{i1} = -c_1 R_{11},$$

(20') can be rewritten as

$$(21) \quad \frac{\partial W}{\partial T_1} T_1 = \frac{1}{2} c_1^2 R_{11} t_1 \left[ t_1 - 2 \sum_{i=2}^n (c_i R_{i1} / -c_1 R_{11}) t_i \right].$$

Thus,  $t_1$  has to be compared with a weighted average of the tax rates on other commodities. Even though we cannot measure the  $R_{i1}$ , so as to know the precise weights to apply, in many cases it is possible to set reasonable limits within which the true weighting pattern is likely to lie. We are likely to have a good idea of which, if any, of goods  $X_2$  to  $X_n$  are very close substitutes or complements to good  $X_1$ . After making allowance for the plausible degree of substitution or complementarity here, we are not likely to go far wrong if we assume that the remaining commodities are remote, "general" substitutes for  $X_1$ . Thus the procedure would be first to estimate  $-c_1 R_{11}$ ; then to estimate  $c_2 R_{21}$  and  $c_3 R_{31}$ , say, if goods 2 and 3 were particularly close substitutes or complements to good one; and finally to distribute the remaining total weights ( $-c_1 R_{11} - c_2 R_{21} - c_3 R_{31}$ ) to commodities  $X_4$  to  $X_n$ , say, in proportion to their relative importance in the national income. Obviously this procedure is not exact, but it is unlikely to lead to a result that is of an erroneous order of magnitude.

B. We now attempt to take account of distortions other than taxes. These can be treated as "autonomous" taxes or subsidies. If a monopoly

is present in industry  $i$ , which prices its products at 20 percent above marginal cost, it is as if a 20 percent tax existed on the product of industry  $i$ , or, perhaps, a 40 percent tax on the value added by labor and by capital in industry  $i$ . If activity  $j$  has positive external effects, leading to an excess of 10 percent of social benefit over marginal cost (at the margin) it is once again as if a tax of 10 percent existed on the value produced in industry  $j$ , or of an appropriately greater percentage on the value added in industry  $j$ . Correspondingly, if an industry's product has negative external effects, it is as if a subsidy existed on the value produced or the value added in that industry (i.e., the economy, by itself, tends to produce too much of that industry's product).

To see how these other distortions would be taken into account, assume that a monopoly exists in industry 1, such that price is  $(1+m_1)$  times marginal cost. Suppose, moreover, that a tax of  $t'_1$  percent exists (or is contemplated) on this product. To take account of the combined effect of the monopoly and the tax, we would simply set  $t_1 = [(1+m_1)/(1-t'_1)] - 1$ , and then use this value for  $t_1$  in (7).

It seems to me that most distortions other than taxes can be taken into account in the way just indicated. There is no intrinsic difficulty, however, in dealing with more complicated cases in which the percentage excess of social value over marginal cost is a function of output rather than a constant. Cases in which the external effects of an industry or activity are independent of its output or level, and depend only on the existence of the industry, need not be dealt with within the framework of this analysis. If the industry or activity is to exist in all situations being compared, external effects of this sort will be equal in all such situations. If, on the other hand, one contemplates eliminating an industry with a given negative external effect, one can calculate by an analysis of the type used in this paper what would be the efficiency-cost of a tax which was just barely prohibitive of the activities of the industry, and see whether this cost outweighed the negative external effect or not.

C. We now turn to a problem which was consciously avoided in Section II. There we maintained the assumption that the capital stock was given. Now we must investigate the possibilities of eliminating this restrictive assumption.

In the first place, we can recognize that, for the analysis of Section II we do not need to assume that the capital stock remains fixed through time. Both population and capital stock can change through time, and the analysis of Section II can be modified to take account of these changes, so long as the changes (in population and in capital stock) are not dependent on tax rates and other distortions. The difficulties appear when we try to allow for the effects of changes in tax rates, etc., on the level of capital stock (and/or population).

Particularly since I have no really satisfactory solution to the prob-

lem posed, I am inclined to defend the assumption that the level of capital stock is reasonably independent of tax rate changes (at least of the sorts of tax rate changes that we have observed in the past). Here I rely on the secular constancy of the rate of net saving in the United States, in the face of substantial savings in the rate of return and in the face of significant alterations in the tax structure. I would not expect, given this historical experience, that the neglect of an effect of taxation upon savings would introduce large errors into the measures derived in Section II.

Obviously, however, this answer, though perhaps adequate for many practical applications, really begs the fundamental question. As I see it, there are three main roads to a solution.

1. One could attempt to extend the "models" of Section II to many time periods, building in all of the relevant dynamics. This, I think, would be scientifically the most satisfying approach to take. However, I am afraid that this approach is likely to complicate the analysis to the point where it will be hard to apply it to real-world problems. Nonetheless, I feel that this is a line worth pursuing.

2. One could attempt to separate the "comparative static" from the "dynamic" costs of alternative tax set-ups. Suppose that changing from tax vector  $T$  to tax vector  $T'$  leads to a change in the rate of saving from  $s$  to  $s'$ . We could measure the change in welfare due to the change in taxes first on the assumption that the rate of saving was unaffected, and then attempt to measure the additional cost or benefit associated with the change in the rate of saving. This approach has a particular appeal because, given the assumption that the net rate of return to capital is equalized in all uses, it is reasonable to assume the rate of saving depends only on the level of real income and the net rate of return.

One can go quite some distance with this approach without greatly complicating the analysis. The present value to the saver of a dollar of saving at the margin is  $\$1.00 = \rho(1-t)/r$ , where  $\rho$  is the social rate of marginal net productivity of capital and  $r$  (which at least in uncomplicated situations should equal  $\rho(1-t)$ ) is the after-tax rate of discount which the individual uses to obtain present values, and  $t$  is the expected future rate of tax on income from saving. The present value of the social yield of capital is simply  $\rho/r$ , so that a dollar's worth of savings should have a social value of  $\$1.00/(1-t)$ . The change in welfare due to the difference in this year's savings stemming from a tax rate of  $t$  rather than a tax rate of zero would then be  $\frac{1}{2}t\Delta s/(1-t)$ , where  $\Delta s$  is the tax-induced change in the amount of this year's savings. If we call this expression  $\Delta_2 W$ , and expression (19), say,  $\Delta_1 W$ , we can express  $\Delta W$  as  $\Delta_1 W + \Delta_2 W$ .  $\Delta_1 W$  expresses the cost this year of misallocating the resources that

would be present this year if the rate of savings were unaffected by tax changes.  $\Delta_s W$  measures the present value of the future benefit foregone because the economy—for tax reasons—did not save “enough” this year. One could correspondingly estimate the  $\Delta W$  stemming from a particular tax structure for a series of future years, and estimate the present value of the future stream of welfare costs associated with that tax structure.

The principal difficulty with approach number 2 is, I believe, that it requires the assumption that  $\rho$  will remain constant in the future. The approach could of course be modified so as to impose a particular non-constant time-path for  $\rho$  in the future, but the basic difficulty remains that the model does not itself tell us what that time-path should be. As a practical matter, however, I believe that changes in the marginal net productivity of capital are likely to be sufficiently slow so that the assumption of constancy will not introduce serious errors in the estimation of  $\Delta W$ .

3. One could attempt to incorporate tax-induced changes in capital stock directly into the analysis. This approach requires two changes in the analysis of cases D and E of Section II. First the assumptions that

$$\sum_i G_{ij} = 0 \quad \text{and} \quad \sum_i H_{ij} = 0$$

must be abandoned; and second, we must eliminate the assumption of the neutrality of any tax striking equally the income from capital in all uses. In effect this means that the only neutral tax treatment of the income from capital would be not to tax it at all.<sup>3</sup> These two adjustments could easily be incorporated into the framework developed in Section II. One additional step would also be necessary. Since the savings-effects of a tax change are likely to go on indefinitely, one would have to decide on the specific time period over which one was measuring the effect of tax changes on the capital stock. This would enable one in principle to deal with specific values for

$$\sum_i G_{ij} \quad \text{and} \quad \sum_i H_{ij},$$

whereas otherwise these values could be almost anything, depending on the time period over which the reactions were being measured. This last requirement—of measurement over a specific time period—is to my mind the most serious disadvantage of approach number 3.

I shall not go into more detail here on the possible merits and disadvantages of the three approaches to the savings problem that I have

<sup>3</sup> Though I find the nontaxation of income from capital repugnant as a policy prescription there is no doubt that even proportional income taxation is nonneutral in respect of the decision to save. The social yield of saving is the gross of tax return to capital, while the private yield is net of tax.

suggested. This problem is, as I have indicated, the most serious "open end" in the analysis of Section II, and I hope that further work in the field, following one or more of the approaches outlined above, will help close this important gap.

## REFERENCES

1. W. J. Corlett and D. C. Hague, "Complementarity and the Excess Burden of Taxation," *Rev. of Econ. Studies*, 1953, pp. 21-30.
2. Gerard Debreu, "The Coefficient of Resource Utilization," *Econometrica*, July, 1951, pp. 273-92.
3. J. R. Hicks, *Value and Capital* (2d ed.; Oxford: Clarendon Press, 1946), esp. pp. 330-33.
4. Harold Hotelling, "The General Welfare in Relation to Problems of Taxation and of Railway and Utility Rates," *Econometrica*, July, 1938.
5. Harry G. Johnson, "The Cost of Protection and the Scientific Tariff," *J.P.E.*, Aug., 1960, pp. 327 ff.
6. Harry G. Johnson, "The Economic Theory of Customs Unions," in *Money, Trade and Economic Growth* (London: George Allen and Unwin, 1962), pp. 48 ff.
7. R. G. Lipsey and K. Lancaster, "The General Theory of Second Best," *Rev. of Econ. Studies*, XXIV, No. 63. (1956-57), pp. 11-32.
8. I. M. D. Little, "Direct versus Indirect Taxes," *Econ. J.*, Sept., 1951.
9. L. W. McKenzie, "Ideal Output and the Interdependence of Firms," *Econ. J.*, Dec., 1951, pp. 785-803.
10. J. E. Meade, *Trade and Welfare*, Vol. II (Mathematical Supplement [London: Oxford Univ. Press, 1955]).
11. Jacob Viner, *The Customs Union Issue* (Carnegie Endowment for Int. Peace, 1950).

## EFFICIENT ALLOCATION OF CAPITAL IN AN UNCERTAIN WORLD

By JACK HIRSHLEIFER  
*University of California, Los Angeles*

In a world of certainty, efficient allocation of capital would be evidenced by equality of the yield (value of marginal product per dollar of capital value) for all forms of investment, waiving certain complications in comparing differing time-shapes of yield streams [13] [2]. And, since in such a world expectations would always be borne out, the equality of yields would be true both prospectively and retrospectively.

When we turn to our actual world of uncertainty, we find that the yields realized on alternative forms of investment differ drastically. In general, there have been two main types of response to this evidence on the part of theorists. According to one school, the evidence indicates "the capital market to be imperfect, to be rife with rationing, ignorance, differential tax treatments, reluctance to finance investment from external funds, slow adjustment processes, etc., which destroy the normative significance of actual rates found in the market" [5, p. 503]. As indicated by the quotation, the positive conclusion about the malfunctioning of capital markets leads directly to a normative inference about the nonrelevance of market interest rates—most immediately, with reference to the controversial question of the appropriate rate of discount for use in evaluating government investment projects not subject to the market test. The other school of thought maintains that the divergences of observed yields conceal an underlying harmony of the capital markets (see, for example, [24]). This view is basically a programmatic hypothesis: those who hold it feel that the search for a consistent structure amidst the seeming confusion of observed yields will ultimately be rewarded. I shall pursue this line of thought here, developing and commenting on two proposed theoretical formulations explaining yield divergences—both turning upon investor attitudes toward risk and uncertainty.

We must note first, however, that in a world of uncertainty, the equilibrating market forces can only work on the prospective returns to investment. There are two possible sources of difficulty here: First, prospects or anticipations are not ordinarily observable and, second, in a world of uncertainty the returns anticipated are multivalued (usually

expressed as a probability distribution). In what follows, I shall develop and compare alternative formulations of the investor's attitude toward multivalued returns; the question of how anticipations about such returns may be made operational for the purpose of direct empirical testing will not be considered here.

One general analytical consideration warrants emphasizing for its bearing upon the examination of alternative theoretical formulations below. Any economic theory of choice should contain, for formal completeness, certain elements: (1) Economic agents (individuals, firms, governments), and objects of choice (ultimately, individuals' consumption); (2) for each agent, preference functions among the objects of choice; (3) again for each agent, endowments of choice-objects and opportunity sets showing the possible transformations of endowments into alternative combinations of desired objects; and (4) conservation relations which constrain the net sum of individual decisions so as to be consistent with social totals of choice-objects available.

### I. *Multivalued Returns— $\mu$ , $\sigma$ Preferences*

The most familiar formulation of investor attitudes toward the multivalued returns of an uncertain world takes for objects of choice the mathematical expectation  $\mu$  and the standard deviation  $\sigma$  of the (subjective) probability distribution of returns— $\sigma$  being interpreted as the measure of the "riskiness" of the investment in question [7] [11] [19] [23] [18] [6].<sup>1</sup> In accordance with the common beliefs of observers of financial markets, and in order to explain the observed phenomenon of diversification of assets, it is usually postulated that investors desire low values for  $\sigma$ —i.e., they are risk-avoiders—and that they show decreasing willingness to accept high  $\mu$  to counterbalance high  $\sigma$  (as  $\sigma$  and  $\mu$  both increase). These assumptions dictate a preference function like that illustrated in Figure 1.

In constructing the opportunity set,  $\mu$ ,  $\sigma$ -theorists have concentrated upon the problem of portfolios; i.e., holdings of financial instruments. Little or no attention has been paid to productive investments. Also, the usual portfolio analysis keeps constant the amount of current investment and concerns itself only with the distribution of that amount over the available securities. Neither restriction is really essential: the same approach could be extended to include productive investments in addition to a financial portfolio, and a simultaneous solution could be provided for the scale of investment, together with the choice of securities held, by incorporating some kind of time-preference model.

In the usual portfolio analysis, the opportunity set is the area bor-

<sup>1</sup> Markowitz goes beyond the other works cited in examining alternative possible statistical measures of "riskiness."

dered by an efficient frontier showing (as in Figure 1) minimum  $\sigma$  attainable for each possible value of  $\mu$ . Since holding a combination of securities whose returns are less than perfectly correlated tends to lower  $\sigma$  relative to  $\mu$  for the overall portfolio, one-security portfolios do not in general appear on the efficient frontier.<sup>2</sup> The curvature shown for the efficient frontier—opposite to<sup>3</sup> that of the  $(\mu, \sigma)$  indifference curves—follows also from the covariance effect, since moving to higher values of portfolio  $\mu$  progressively reduces the number of securities that can be held in combination so as to lower  $\sigma$ .<sup>4</sup>

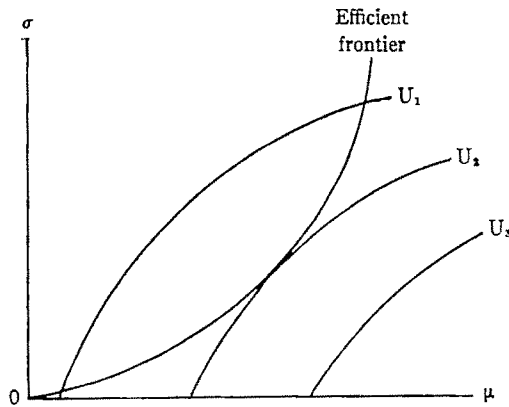


FIGURE 1  
 $\mu, \sigma$ -PREFERENCES AND EFFICIENT FRONTIER

The approach in terms of  $\mu, \sigma$ -preferences has the advantages of plausibility, and a fairly direct connection between theoretical and measurable magnitudes (means, variances, covariances and so forth are all observable, though only retrospectively). A number of objections can be raised, however, of which only two will be mentioned here—leaving aside for the moment the complex question of the consistency of observed market behavior with the risk-aversion assumption.

1. It has been pointed out that it is not possible to derive  $\mu, \sigma$  preference functions from the Neumann-Morgenstern postulates of rational choice except with the aid of arbitrary restrictions upon the subjective probability distributions or utility-of-wealth functions [23] [3].<sup>4</sup>

2. Another objection, less widely appreciated, is that current formu-

<sup>2</sup> However, the security with highest  $\mu$  (whatever its  $\sigma$ ) is on the efficient frontier. Also, if there are riskless securities, the riskless security of highest  $\mu$  must also constitute an efficient one-security portfolio.

<sup>3</sup> If, however, the available portfolios are all combinations of a single riskless security ( $\mu = \mu_0, \sigma = 0$ ) and a risky one ( $\mu = \mu_1, \sigma = \sigma_1$ , where  $\mu_1 > \mu_0, \sigma_1 > 0$ ), the efficient frontier reduces to a line (see [23]).

<sup>4</sup> The special assumptions are that the probability distribution must be two-parameter, or alternatively that the utility-of-wealth function be quadratic. The latter is especially objectionable, as it requires the marginal utility of wealth to become negative beyond a certain point.



lations in portfolio analysis do not meet the requirements for a complete choice-theoretic system. Granted that individuals have  $\mu$ ,  $\sigma$ -preferences, it remains to be shown how the resulting decisions of individuals in the market interact to determine security prices that permit the existing stocks of securities to be exactly held in market equilibrium.<sup>5</sup> Furthermore, securities are artificial commodities; still further steps are necessary to show how and why they are generated. Such an analysis must go back to the forces determining the balance between  $\mu$  and  $\sigma$  in real investments undertaken,<sup>6</sup> the principles on which owners of risky income streams partition  $\mu$  and  $\sigma$  to make up the various classes of securities that represent claims to such streams, and the conservation relations that apply to the social totals of  $\mu$  and  $\sigma$  in the economy. In short, what is lacking is a theory of the market risk-premium.

## II. *Multivalued Returns—Time-State Preferences*

A less familiar, but (I feel) ultimately more satisfactory approach to decision under uncertainty has been pioneered by Arrow and Debreu [1]. In this formulation the objects of choice are not derivative statistical measures of the probability distribution of consumption opportunities but rather the contingent consumption claims themselves set out in extensive form. Combining Fisher's treatment of time-preference decisions under certainty with Arrow's conception of choice objects under uncertainty permits the formulation of a generalized theory in which the objects of choice are consumption opportunities as of alternative dates and specified states of the world—or, we shall say, alternative time-state claims. If the present and future are strictly determined, the objects of choice reduce simply to consumption opportunities of differing dates, as in Fisher's riskless time-preference system [8].<sup>7</sup> It is evident that Fisher's time-claims to consumption generalize immediately to time-state claims, that under pure exchange the social total of each class of choice-objects must be conserved, and that analogous formulations for productive opportunities can be set down. The chief novel element is that, for state-preference at a given time, the Neumann-Morgenstern theorems permit a special formulation of the preference function.

In the interests of simplicity, in what follows it will be assumed that there is only one present state; i.e., there is certainty as to the present (time 0). The future is represented by a point in time (time 1) in which there are two alternative states (state *a* or state *b*). The two states may be thought of as war versus peace, or prosperity versus depression. The

<sup>5</sup> Some recent work by William Sharpe makes important progress in this direction [22].

<sup>6</sup> An important element in such decisions must be the physical "productivity of risk" [14].

<sup>7</sup> Fisher, who may have been the first to suggest a  $\mu$ ,  $\sigma$  approach to risk-choices, apparently never employed a time-state formulation for uncertainty.

fundamental objects of choice, contingent consumption opportunities, may be denoted  $c_0$ ,  $c_{1a}$ , and  $c_{1b}$ . Whereas in the deterministic time-preference case there is only one independent dimension of choice (between  $c_0$  and  $c_1$ ), here there are two: a synchronous balance between the contingent claims  $c_{1a}$  and  $c_{1b}$ , and a present-future balance (for the latter, it suffices to show the exchange between  $c_0$  and  $c_{1b}$ , since the proportions of  $c_{1a}$  and  $c_{1b}$  are determined in the synchronous exchange).<sup>8</sup>

In Figure 2a the 45° "certainty line" through the origin represents points along which  $c_{1a} = c_{1b}$ . Convexity follows from the empirical observation of "nonspecialization" of risky choice. Almost no one is so reckless as to prefer—if, say,  $p_a = p_b = \frac{1}{2}$ —the prospect  $(c_{1a}, c_{1b}) = (1000, 0)$  to a prospect like  $(500, 500)$ . (It must be understood that  $c_{1b} = 0$  does

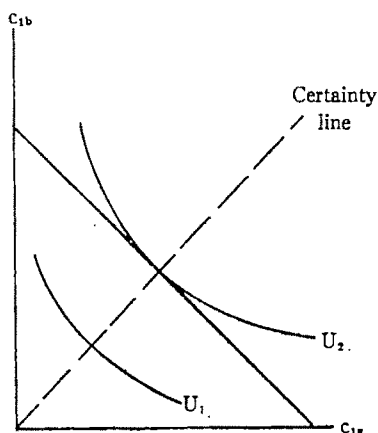


FIGURE 2a  
CONTEMPORANEOUS STATE-PREFERENCES AND  
FINANCIAL OPPORTUNITIES  
( $t = 1$ ,  $p_a = p_b = \frac{1}{2}$ )

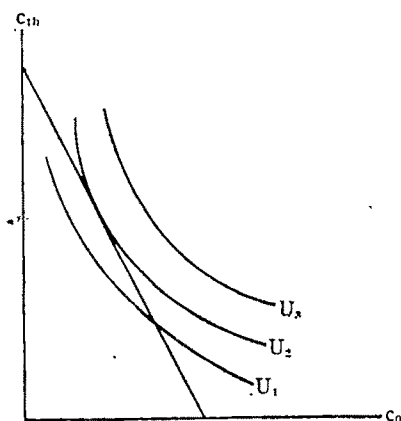


FIGURE 2b  
TIME-AND-STATE PREFERENCES AND  
FINANCIAL OPPORTUNITIES

not mean merely the unfavorable outcome of a lottery, representing a possibly minor decrease in overall wealth but rather a zero-consumption situation—absolute impoverishment should state  $b$  occur.) Thus, even a very mild degree of conservatism requires convex utility isoquants.

If we now introduce, for consumption at time 1, a Neumann-Morgenstern utility-of-income function  $v(c_1)$  (see [9]), interesting results follow. It can be shown that, if the function  $v(c_1)$  does not itself change whether state  $a$  or  $b$  obtains, everywhere-convex indifference curves

<sup>8</sup> For some purposes, it is more convenient to distinguish the synchronous (risky) exchange between  $c_{1a}$  and  $c_{1b}$  from the riskless time-exchange between  $c_0$  and a certain claim to  $c_1$  (see paragraph below).

require a "concave" utility-of-income function—that is, one characterized by diminishing marginal utility [15]. Thus, Neumann-Morgenstern risk-aversion follows from observed nonspecialization among contingencies.

In a pure-exchange situation, the individual's opportunities are limited to converting his endowment time-state distribution,  $Y$ , into a consumption time-state distribution,  $C$ , subject to a wealth-constraint  $W$ .

$$W = y_0 + y_{1a}/(1 + r_{1a}) + y_{1b}/(1 + r_{1b}) = c_0 + c_{1a}/(1 + r_{1a}) + c_{1b}/(1 + r_{1b}).$$

Here  $r_{1a}$  is the time-and-probability discount rate for contingent claims of date 1 and state  $a$ . We may also write the wealth constraint as  $W = P_0 c_0 + P_{1a} c_{1a} + P_{1b} c_{1b}$  and, with  $c_0$  as numeraire,  $P_0 = 1$ ,  $P_{1a} = 1/(1 + r_{1a})$ , and  $P_{1b} = 1/(1 + r_{1b})$ .

For the contemporaneous exchange between  $c_{1a}$  and  $c_{1b}$ , it can be shown that if the price ratio  $P_{1b}/P_{1a}$  is equal to the probability ratio  $p_b/p_a$ , the optimum (for a consumer whose Neumann-Morgenstern  $v(c)$  function is independent of state) must be along the certainty line. This corresponds to the well-known theorem that, if  $v(c)$  is concave, a fair gamble—and, a fortiori, an adverse gamble—will not be accepted. Thus, an alternative explanation would have to be sought for the phenomenon of Las Vegas. (It may be noted here that the customary betting pattern at Las Vegas—repeated commitment of small stakes—makes the overall outcome close to a certainty; thus suggesting that participants in this game are not seriously attempting to change their wealth levels. If they were so attempting, they would hazard large fractions of their wealth on a single turn of a card or roll of a wheel. What they are really doing, one is forced to conclude, is paying a modest fee for a certain thrill or excitement under highly controlled conditions involving minimal real risk.)

A somewhat novel idea to which the time-state approach leads, however, is that the Neumann-Morgenstern function itself may not be independent of state—since different states are not mere gamble outcomes like Black or Red at roulette, but rather may represent objective differences in one's external or internal context for choice. If, for example, the states are personal health versus illness, even an unfair "gamble" might be accepted, provided it pays off disproportionately in the state in which income is more urgently needed; thus, purchase of health insurance, while an adverse gamble, remains essentially conservative.<sup>9</sup> Convexity of the indifference curves would ordinarily dictate an interior solution, so that only moderate gambles will be ac-

<sup>9</sup> The variation of the  $v(c)$  function might be attributed to the presence of positive or negative, "nonpecuniary income" in one of the states envisaged.

cepted.<sup>10</sup> This generalization of risk-aversion might be termed "conservative behavior."

To complete the choice-theoretic system under pure exchange, conservation equations must be incorporated which require that the entire social endowment, in each time-state, be exactly distributed among economic agents. If productive opportunities are introduced, the goal of production for each agent is wealth-maximization—where wealth is the sum of contingent income claims discounted, for time and state, by factors like  $(1+r_{1a})$ . One problem is to explain why the capital markets ordinarily deal not in the "natural" units of time-state claims like  $c_{1a}$  or  $c_{1b}$ , but in securities representing complex packages of such claims. The reason seems to be the huge number of conceivable contingencies; some aggregation of claims is required. The main aggregation procedure observed in the market is the separation of the claims resulting from investments into one package that is relatively certain or "senior" (bonds), and another covering the residuals (stocks). Treating each class of securities as a bundle of time-state claims, it is possible to show (under certain idealizing assumptions, most importantly zero taxation and absence of liquidity premiums) that the sum of values of alternative security combinations must be a constant—equal to the sum of values of the underlying physical time-state claims [15]. This is, of course, the Modigliani-Miller theorem [21, p. 268].

### III. *Harmony of Market Yields*

Does the foregoing formulation of investment theory for an uncertain world square with observed patterns of market yields? In particular, does it go some distance toward showing that observed divergences represent, when properly interpreted, a harmonious structure of yields? Only a few remarks on the evidence can be put forward here.

1. The great diversity of experienced yields, both cross-section and over time. This, of course, is exactly what one would expect as the economy threads its way among multiple possibilities over time [4]; different securities, although equally valued by the market *ex ante*, will be differentially affected *ex post* as some one of the set of mutually exclusive contingencies eventuates at each date.

2. Bond yields are less than stock yields (averaged over time) [17]. This is consistent with risk-aversion (postulated on the different ground of nonspecialization among states). It may be worth noting that returns from holding high-grade bonds are better, from a risk-avoiding point of view, than a merely stable yield. A constant-dollar bond return will

<sup>10</sup> Friedman and Savage [9], however, maintain that observed behavior shows aversion to moderate gambles as compared with either highly secure or highly risky investments. Their illustrations can, I would maintain, be explained more satisfactorily on other grounds.

fluctuate upward in real terms during deep depression—just when income is most urgently needed.

3. Normally rising term-structure pattern.<sup>11</sup> This is usually interpreted as evidencing liquidity-preference, and in the Hicksian formulation is explained in terms of risk-avoiding behavior [12, pp. 145–47]. The pattern is thus consistent with convex time-and-state preferences. However, a time-state analysis employing a complete choice-theoretic system (too long to reproduce here) indicates that persistence of the long-short differential involves, in addition to risk-aversion, certain technological phenomena: positive net productivity of lengthening the “period of production,” and limited reversibility of physical investments in the event of an adverse contingency (with consequent need for current income).

4. High yields on human investment. An implication of the foregoing is that investment in oneself—being exceptionally irreversible—should on the margin show high returns, which appears to be borne out by the evidence.<sup>12</sup>

#### IV. *Concluding Comment*

As indicated earlier, the harmonistic hypothesis is essentially a programmatic one. However, the formulation of investment choice under uncertainty in terms of time-state preferences, with the assumption of risk-aversion (or rather, the slightly generalized assumption of conservative behavior) does seem to promise progress toward harmonizing the bewildering diversity of market yields. I will conclude with the remark that validation of the harmonistic hypothesis would suggest, for the policy question alluded to earlier concerning the appropriate discount rate for government investment projects, the use of market yield rates comparably placed in the rate structure—taking account of risk, term, illiquidity, etc., or ultimately of the alternative time-states in which the investment income accrues. This would indicate considerably tougher standards than the current practice of employing the government's risk-free borrowing rate.

<sup>11</sup> Meiselman [20] has shown that the term structure can, consistently with the data, be interpreted as reflecting changing expectations over time. But Kessel, in a still-unpublished N.B.E.R. study [16], has demonstrated that there remains a bias in the form of a normal excess of long rates over short. See also [25].

<sup>12</sup> W. L. Hansen shows marginal private rates of return in schooling clustering mainly between 10 percent and 20 per cent [10]. There are, it must be admitted, many difficulties in the data and interpretation.

#### REFERENCES

1. K. J. Arrow, “Le Role des Valeurs Boursieres pour la Repartition la Meilleure des Risques,” *International Colloquium on Econometrics* (1952).
2. Martin J. Bailey, “Formal Criteria for Investment Decisions,” *J.P.E.*, Oct., 1959.
3. Karl Borch, “A Note on Utility and Attitudes to Risk,” *Management Science*, July, 1963.
4. Gerard Debreu, “Une Economique de L'Incertain,” *Economie Appliquée*, 1960.

5. Otto Eckstein, in National Bureau of Economic Research, *Public Finances: Needs, Sources, and Utilization* (Princeton Univ. Press, 1961).
6. D. E. Farrar, *The Investment Decision Under Uncertainty* (Prentice-Hall, 1962).
7. Irving Fisher, *The Nature of Capital and Income* (Macmillan, 1912), Chap. 16.
8. ———, *The Theory of Interest* (Macmillan, 1930; reprinted, Augustus M. Kelley, 1955).
9. M. Friedman and L. J. Savage, "The Utility Analysis of Choices Involving Risk," *J.P.E.*, 1948. Reprinted, *Readings in Price Theory* (Irwin, 1952).
10. W. L. Hansen, "Total and Private Rates of Return to Investment in Schooling," *J.P.E.*, April, 1963.
11. J. R. Hicks, "A Suggestion for Simplifying the Theory of Money," *Economica*, Feb., 1935.
12. ———, *Value and Capital* (2nd ed., Oxford Univ. Press, 1946).
13. J. Hirshleifer, "On the Theory of Optimal Investment Decision," *J.P.E.*, Aug., 1958. Reprinted in E. Solomon, ed., *The Management of Corporate Capital* (Free Press, 1959).
14. ———, "Risk, the Discount Rate, and Investment Decisions," *A.E.R.*, May, 1961.
15. ———, "Investment Decision Under Uncertainty" (1963).
16. Reuben A. Kessel, "The Cyclical Behavior of the Term Structure of Interest Rates" (1962).
17. Henry A. Latane, "Portfolio Balance—The Demand for Money, Bonds, and Stock," *Southern Econ. J.*, Oct., 1962.
18. Harry Markowitz, *Portfolio Selection* (Wiley, 1959).
19. J. Marschak, "Money and the Theory of Assets," *Econometrica*, Oct., 1938.
20. David Meiselman, *The Term Structure of Interest Rates* (Prentice-Hall, 1962).
21. F. Modigliani and M. H. Miller, "The Cost of Capital, Corporation Finance and the Theory of Investment," *A.E.R.*, June, 1958.
22. W. F. Sharpe, "Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk" (1963).
23. James Tobin, "Liquidity Preference as Behavior Towards Risk," *Rev. of Econ. Studies*, Feb., 1958.
24. ———, "Money, Capital, and Other Stores of Value," *A.E.R.*, May, 1961.
25. John H. Wood, "Expectations, Errors, and the Term Structure of Interest Rates," *J.P.E.*, April, 1963.

## DISCUSSION

DALE W. JORGENSEN: Once upon a time not so long ago it was the fashion to present the theory of general economic equilibrium as a kind of fable. The stage was set by the listing of numerous "unrealistic" hypotheses—the whole world is a single point in space, all activity takes place at an instant in time, all events occur with perfect certainty. Every budding economist was trained in the art of constructing plausible "counterexamples" to these hypotheses. Judging this educational process by its results, the student has learned his lesson when he approaches each applied problem *de novo*. To justify this procedure, the well-trained economist begins his discourse with a list of the unrealistic hypotheses of theory. He then goes on to state: in the real world of uncertainty, where activity takes place through time and at different locations, and so on, and so on, we must proceed on an *ad hoc* basis, setting aside any considerations of abstract theory.

From a purely pedagogical point of view, the most important discovery emanating from the Arrow-Debreu theory of general equilibrium is that all of the unrealistic assumptions of the old-style economic theorist as listed above can be set aside. To incorporate the complications associated with location, time, and uncertainty, no logical extension of the theory is required. All that is necessary is to reinterpret the notion of a commodity. Instead of taking as examples such commodities as corn and iron, we must take corn delivered at Chicago next month if war does not break out, and so on. This discovery is so simple and so fundamental that we will be extremely fortunate if it finds its way into the conventional discourse of economists within the next generation.

In this connection it is a highly auspicious development for economic theory that Professor Hirshleifer has turned to consideration of the problem of uncertainty. He has already revived and extended the nearly forgotten contributions of Irving Fisher to intertemporal economics. Within the present context, Fisher's main contribution was to reinterpret the notion of commodity to include commodities delivered at various points in time. Fisher demonstrated that given such a reinterpretation, the intertemporal theory of individual behavior and of general equilibrium is formally identical to so-called "static" general equilibrium theory. This masterful conception is not widely appreciated even now, as is evidenced by the feeling of both mystery and awe engendered in most students of economics by utterance of the phrase "capital theory." But a good part of what appreciation and understanding of Fisher's ideas there is among economists must be attributed to Professor Hirshleifer.

In the present paper Hirshleifer has provided an exposition of Kenneth Arrow's highly important contribution to the theory of economic behavior under uncertainty. Arrow's contribution is very much in the spirit of Fisher's; the notion of commodity is reinterpreted to include commodities delivered

under various contingencies. Just as Fisher takes a commodity like corn at two different points in time as two commodities, Arrow takes a commodity at two different states of the world (say, war or peace) as two commodities. As Gerard Debreu has shown, a similar reinterpretation of the notion of commodity provides an extension of the theory to economic activity at several different points in space. Of course, it is possible to combine all three attributes of a given commodity—location, time, and state of the world—to provide a logically complete theory of general equilibrium. This has been carried out by Debreu in Chapter 7 of his *Theory of Value*. Debreu's exposition, though concise, is not highly technical and may be recommended as a useful supplement to Hirshleifer's present paper.

Turning to Professor Harberger's contribution on the measurement of economic waste, we find assurance that the theory in which the whole world is a single point in space, all activity takes place at an instant of time, and all events occur with perfect certainty is accepted in some parts of the economics profession as something other than a fable. Confronted with a choice between no theory at all as a basis for applied work and old-style general equilibrium theory, even in its Marshallian form, one must rise to applaud the valor of those who, like Harberger, elect theory. But from the point of view of Fisher, Arrow, and Debreu, the choice between no theory and old-style theory involves a false dichotomy. The choice to be made at the present time is between the modern theory of general equilibrium and the ancient fable which most of us have been brought up to disbelieve. Harberger suggests in effect that we should elect the fable.

Harberger's point of view deserves a hearing in any discussion of "positive" economics, if only on the grounds that any fantastic presupposition may lead to interesting empirical results. But Harberger's view has implications for the "normative" side of economics which he does not make fully explicit. In assuming that the supply of "capital" is fixed and independent of all distortions now and forever, Harberger goes a step beyond the usual assumption of no distortions except where the waste-measurer is currently measuring waste. He assumes that the distortions themselves affect the situation only in the markets for commodities to be delivered in the current period. It is implicit in this assumption that these same distortions do not affect the markets for commodities to be delivered in all future periods.

The palpable absurdity of Harberger's assumption of a fixed supply of capital over time (not necessarily the same at each point in time) may or may not prevent its becoming a useful working hypothesis in an empirical study. In the economics of welfare, this assumption is not merely absurd; it flies in the face of the main line of recent developments in the welfare economics of the second best. By assuming away distortions in markets for commodities for future delivery, Harberger leaves open the very real possibility that policy recommendations resulting from his measurements may do infinitely more harm in the markets for future commodities than the small amount of measurable good they do in the markets for present commodities. This consideration is of paramount importance for almost all of the problems for which Harberger wants "practical" results: depletion and capital gains provisions,



the corporation income tax, taxation of capital versus taxation of other factors of production, taxation of real property.

All of the foregoing remarks are directed to Harberger's underlying assumption of a one-period world. His attempt to fill the logical gap in the analysis introduced by this assumption, insofar as the attempt is a serious one, is confined to different ways of getting the assumption of fixed capital stock in through the back door; for example, through the alleged inelasticity of savings or productivity of capital schedules. Similar remarks could be addressed to Harberger's assumption of perfect certainty. Professor Hirschleifer has demonstrated the fruitfulness of permitting uncertainty to be taken into account in the evaluation of government investment projects.

While endorsing Harberger's motives in attempting to use an economic fable for welfare analysis of what he calls real world problems, we must withhold endorsement of results derived from this fable. From the point of view of measurement alone, Harberger's results may go awry by several orders of magnitude. From the policy point of view the results may point in just the wrong direction from that which Harberger has in mind. To the four possible reasons for unpopularity of the loss-measurement game listed by Harberger, we may add a fifth:

To measure economic waste at a point in time, it is necessary not only to consider markets for " $n$  commodities" like corn and iron, but markets for an indefinitely large number of commodities like corn delivered at Chicago next month if war does not break out. Some of these markets actually exist; for example, one may observe the price today of a dollar tomorrow if the federal government does not suspend payment of its obligations. From the observed data one could, at least in principle, measure the relevant elasticities of demand and supply for the commodity in question. But of the total number of commodities which enter directly into the calculations of such economic actors as investors and savers, the proportion of commodities for which there are active markets or any markets at all is miniscule. Therefore, even for the true believer in the economics of the  $n$ th best, the calculations required for a determination of waste in a world with more than one time period, more than one location, and more than one possible future state cannot be carried out.

**WILLIAM VICKREY:** Harberger has presented us with a set of alternative procedures for measuring inefficiency by means of certain quadratic forms; the availability of alternative formulations which, given the assumptions, yield equivalent results, makes it possible in some cases to get estimates that might seem hopelessly beyond our reach if approached from other perhaps more natural directions. Of course, the results are subject to errors resulting from the neglect of the higher order terms in the expansion of underlying relations, as well as to the usual caveats regarding interpersonal comparisons. Yet as a practical matter, decisions are going to be made, and to neglect altogether such evidence as might be supplied by the methods brought forth by Harberger would be to reject a tool because it does not meet impossible standards of perfection.

Indeed, while choice among alternatives is an everyday matter, it is the

particular province of political economy to choose among methods of choice. In a particular instance, there may be some likelihood that a decision based on a comparison of waste measured by the Harberger methods would be incorrect or questionable on the ground that the gainers and losers are in some way not comparable. But if an overall decision is to be made between accepting such a measure of waste as a criterion to be given considerable weight in the making of decisions, and refusing to give some such measure any consideration in the making of decisions, it seems highly likely that only a small minority, if any, of individuals would, in the event, fare significantly less well when the criterion is used than when it is not. Moreover, if one thinks of this choice among methods of choice as being made *ex ante*, with the field of decisions to which the method of choice is to be applied somewhat vaguely defined and the results somewhat uncertain, particularly as they apply to any one individual, the number of individuals whose well-informed expectations would be impaired by electing to give substantial weight to some such measure of inefficiency in all future decisions is likely to approach the vanishing point. The decision as to methods of choice could thus be reached with virtual unanimity.

The operational significance of Hirshleifer's paper for me lies in its bearing on the proper rate of interest to use in the evaluation of public capital investments, and particularly with respect to how this rate should vary with variations in the riskiness of such investments and in their duration. Hirshleifer rather suggests that market rates for comparable terms, liquidity, and riskiness should be accepted as a first approximation guide. But insofar as riskiness is concerned, the reason risky investments carry an expected return greater than that of secure investments is that in the market, facilities for pooling of risks are imperfect, so that investment by private investors in risky investments, given a limited portfolio and a certain indivisibility in the market, is unavoidably associated with a considerable dispersion of individual incomes. Given a certain risk-aversion on the part of individuals and supposing that this can be interpreted in terms of a Bernouillian utility, maximization of expected utility leads naturally to a bias against risky investments.

On the social level, however, the risk associated with a given public venture is inevitably pooled and averaged over the entire population of the country in some fashion, along with the risks of other projects, and this pooling or averaging of risks for public projects is accomplished without any cost of extra financial transactions. And for almost any project that falls short of the Manhattan project in overall magnitude, the variance in the expected national income associated with a particular project will correspond to such a small fraction of the national income that it would be unlikely that the average utility function of individuals over the relevant ranges would show significant curvature such as to warrant a margin in the interest rate charged over the rate of return on safe bonds.

Insofar as liquidity is concerned, the rationale according to which certain individuals pay a premium for liquidity in terms of low rates of return on short-term securities is associated with the risk that a need for cash earlier than the maturity of longer-term instruments would coincide with a rise in

interest rates. This risk is apparently more prevalent than the inverse one that a surplus of liquidity at an earlier date would coincide with a reduction in interest rates leaving less favorable opportunities for investment. Again, the long-run tendency of short-term rates to run lower than long-term rates, or for long-term rates to exceed an average of the short-term rates over a comparable total term, is related to the circumstance that in markets as we know them, individuals in the first case outweigh those in the second, and not to any inherent and all-pervasive liquidity preference. That arbitrage fails to bring the long-term rate into line with the expected short-term rates over comparable periods is related to the fact that firms that can switch from borrowing long to borrowing short are in general limited in the extent to which they can do this by the effect of such practices on their credit standing. The one entity with a sufficiently firm credit standing to be able to balance the market by switching from borrowing long to borrowing short is of course the government; it is not at all clear why the government should refrain from doing this.

If short-term instruments are not thus supplied by government or otherwise to the extent demanded to balance expected short-term and long-term rates, the general tendency would be for inventory and other short-term investments that can be financed from short-term funds to be carried to a point of smaller marginal returns than long-term investments. The less productive distribution of investment would be called for in order to meet the risk-aversion of individuals faced with uncertainties as to future rates of interest. If the government can balance the market on the basis of its credit standing, however, there would seem to be no warrant for pushing short-term investment to a point where lower rates of return are obtained than from long-run investment. If it is felt necessary, for reasons of monetary stability, to limit the amount of short-term borrowing by the government (a necessity I question), the resulting bias in private investment may have to be tolerated as a price to be paid for monetary stability; but in any case this would not justify insisting on this wasteful bias being fruitlessly extended to government investment. In other words, the rates of return required on short-term or more liquid government investments should not be systematically lower than for less liquid or long-term government investments.

On the other hand, once short-term and long-term rates have been harmonized, in the sense that there is no general expectation that short-term rates will turn out to average either higher or lower than long-term rates over comparable periods, it may be that the short-term interest rate should be consulted even in deciding upon long-term projects. Investment decisions, especially those made by governments, are not always of the "now or never" variety; often they are decisions between making an investment now or reconsidering the matter at a later date. The most promising alternatives may be investing now and investing later, permanent omission of the investment ranking as a rather poor third. In deciding between the alternatives of investing now and investing in the near future, it is the short-term rate of interest that is relevant, whether in its own right or as a reflection of expectations regarding the future long-term rate. In a situation where capital is temporarily

scarce, as might for example be the case immediately after a major disaster, this would be reflected in relatively high short-term interest rates and an expectation that these rates would gradually fall to a more normal level; investments to be considered for immediate adoption should be those with early returns at a relatively high level, or at least those whose postponement would entail a more serious loss of benefits, even though some of these lost benefits might occur only later in time, as with a project with a substantial gestation period. If the current long-term rate for a term equal to the duration of the project is used as a basis for the decision, there is a likelihood that construction may be started on a project because it promises a return higher than the interest cost on its long-term financing, overlooking the even more favorable alternative of planning to begin the project after a delay of a year or so. In effect, the "rent" attributable to the situation making the investment desirable will not be maximized. To be sure, if the postponement alternative is specifically considered on the basis of an expected lower long-term rate of interest, this superior alternative would still be revealed in terms of long-term rates. For my part, however, the analysis in terms of short-term rates seems both conceptually simpler and more likely to call attention forcibly to the better alternatives. Even so, of course, the comparative returns from immediate and deferred construction need to be computed with appropriately distributed depreciation charges.

But while I would thus come to the conclusion that neither liquidity preference nor risk premium (in the sense of a differential between expectation of returns on safe and risky projects) are appropriate factors to be taken into account in figuring the capital cost of public projects, this does not mean that one should take the rate on riskless long-term securities as a standard. Under assumptions of constant degree of underemployment (and without this assumption market rates of interest become totally irrelevant) the resources for public capital investment must come either from consumption or investment. Private investment, at the margin, is subject to very substantial imposts in the form of corporation income tax, property tax, and the like, leading to a strong presumption that where public projects cannot show benefits covering an equivalent for such charges, then if resources for public investment are withdrawn from private investment, a reduction in the overall productivity of investment will take place. On this basis a convenient method of allowing for this factor would be to use a rate of return comparable to a rate gross of tax, which might be of the order of 8 to 10 percent or more rather than the more usual 4 to 6 percent.

On the other hand it is possible to suppose that public investment resources are drawn from current consumption. In this case some form of taxation will usually be involved, and the considerations discussed in Harberger's paper become relevant. The completely neutral tax that he uses as a tool of analysis of course does not exist in stern reality, at least in a form that is capable of much expansion at the margin, although I did hear recently of an attempt in South America to propose a tax on "facultative income"; i.e., that income that one would earn at one's occupation if working a standard number of hours per week. Accordingly, one is more or less compelled to consider

that funds raised for investment out of consumption by taxation are surchargeable with an added cost reflecting the added interference with optimal adjustments caused by the increased tax rates. In principle, one could speak of the "marginal cost of public funds" according to which each dollar of revenue raised by increased taxation would be found to cost the taxpayers \$1.00 plus  $z$  of satisfactions forgone;  $z$  might range from ten cents in a jurisdiction with a well-designed, efficiently administered, and not too heavy tax system, to perhaps fifty cents or even more in a jurisdiction with limited taxing authority, with its revenue resources pressed to the limit, as is the case with many cities. Again, one way of including this element of the marginal cost of public funds would be through a corresponding margin applied to the rate of interest.

Of course as soon as one talks about public investment with resources drawn from consumption one is faced with the question of the relation between market rates of interest and private marginal time preferences; income taxes and other influences may cause market rates to exceed marginal time preference rates especially for those in the higher brackets. But then there is the whole rather philosophical question of how to determine the allocation of resources between the present and the future on a social as distinct from an individual basis.

Alternatively, one may suppose that limitations on total investment by government are more or less imposed by budget notions, rational or irrational, but in this case the appropriate rate of interest for investment decision making is an internal one generated by opportunities within the public sector and divorced almost completely from rates in the private investment market. But in any case government investment decisions should not be based, even in a preliminary evaluation, on rates of interest that include allowances for dispersion of yield expectations or for liquidity, or fail to include a margin in lieu of taxes.

TJALLING C. KOOPMANS: While basically in sympathy with Hirshleifer's arguments in favor of the contingent commodity approach, I should like to point to one argument possibly favoring the  $(\mu, \sigma)$  approach that seems to have been overlooked in the literature.

As pointed out by Hirshleifer, efficient portfolios however defined are, in most cases of interest, also diversified portfolios. The diversification is the more effective in attaining efficiency, the smaller the correlations between anticipated returns of assets that are prominent in the portfolio. But, by the central limit theorem of probability calculus, the sum of a sufficient number of sufficiently independent random variables each of limited variance has a near-normal distribution. The question thus naturally arises whether among the assets available for investment there is sufficient independent variability to cause approximate normality of the distribution of anticipated returns to be a (welcome if unintended) by-product of the quest for an efficient portfolio.

This is obviously an empirical question that cannot be answered solely from conceptual analysis. The pervasiveness of general cyclical fluctuation

in the returns on many stocks might lead one to doubt that an affirmative answer can be given. However, the availability in the capital market of bonds, real estate, and other physical property, together with the possibility of holding negative amounts of some assets by uncovered sale for future delivery, should also be taken into account.

The empirical question raised is important because an affirmative answer would greatly simplify portfolio analysis. The effect of the purchase of an asset on mean and variance of the portfolio would be all that an investor would need to consider.

Finally, even if for the present real world the answer to our question were to be negative, one might conjecture that advances in the control of cyclical fluctuations would, again as a by-product, make the returns on available assets conform more to the premises of an appropriate version of the central limit theorem.

I now come to the discussion of Harberger's paper. Sitting at the table on the podium here I had the privilege of reading a copy of Jorgenson's remarks as he made them. I cannot quite see the contrast he makes as one between an outdated fable embraced by Harberger and—what is it, the revealed truth of the Arrow-Debreu model of competitive equilibrium? It does seem to me, however, that the latter model, and related models by other authors, have set a new precedent and standard of explicitness and completeness in taking into account all the interrelations between variables, to which the premises of the model give rise. This explicitness is sadly lacking in Harberger's paper. To ask assent to the statement that, in the relation

$$X = f(D, G, T)$$

from which his reasoning takes off, effects of changes in tax rates  $T$  on the distribution of income  $D$  can be offset by neutral taxes and transfers, places a heavy burden indeed on the intuition of the reader. One would like to see that relation, as well as the expression (1) for  $\Delta W$ , derived from explicitly stated more basic behavior relations of individuals. Some reason for doubt about Harberger's procedure arises from a study by Debreu of 1954<sup>1</sup> in which the neglected income effects are found to be of the same order of magnitude as the substitution effects on which Harberger concentrates.

PAUL A. SAMUELSON: I agree with Koopmans' remark that Harberger builds an upper floor in a house that needs established foundation. But I defend him against the quaint view of Jorgenson that the (allegedly recent!) discovery of how we can treat corn in Chicago as a different good from corn in Denver or from corn in Chicago tomorrow represents a point against Harberger's theoretical system. If Harberger's  $\Sigma \Delta P \Delta Q$  measures were theoretically valid at a point now, economists have long known they would be theoretically valid in a Fisher world involving future goods. (Of course, workshopmen would find it harder to stuff those "empty boxes.")

<sup>1</sup> G. Debreu, "A Classical Tax-Subsidy Problem," *Econometrica*, Jan., 1954, pp. 14-22. This study still falls short of the goal in that, as its premises clearly state, it does not cover tax changes that result in changes in even a single consumer's list of commodities consumed in positive amounts.

Simple consumer's surplus is known to be exact only in the empirically bizarre case of parallel indifference curves (where it happens to be superfluous). The ancient theory of its approximate validity remains a modern mess; but the generalized theory of revealed preference tells precisely what inferences about individual and group preferences can be derived from various  $(P, Q)$  observations. In the end, the economist can hope only to make certain statements about (1) whether an individual is ordinally better off in situation A than B, (2) whether a specified Bergson social welfare function is better there, or (3) whether certain feasible production-possibility and ordinal-utility-possibility frontiers have shifted outward locally or globally. Waiving technical flaws, I believe that the proposed dollar efficiency measurements are likely to end up merely telling us whether the Hicks-Kaldor criterion is satisfied. This much-debunked criterion, which favors the *status quo* and a Panglossian theory of property, is properly open to the Scitovsky objection (and my later generalization of it).

But even if that were overcome, the whole procedure may remain misleadingly pretentious. I am reminded of the case where a jury must determine which of two stones is heavier. The first stone can allegedly be weighed to a high degree of numerical accuracy (dollar efficiency measures). The second cannot in principle be so weighed (value-judgment decisions). What is the point in measuring the first stone if in the end we must put that exact weight in the balance with the second—particularly if, as is the true case, the two stones have no independent existence and weights? Harberger's valid defense would be to admit he is really trying to make estimates of possibility-frontier shifts. To answer, as he did to Vickrey, that since economists are no good at weighing the second stone they should concentrate on weighing the first is to fall prey to the pathetic fallacy that the universe was created to keep economists busy at something or other.

There is a technical flaw in using the device of holding the percentage distribution of money incomes constant (whatever that is supposed to mean in a world where people differ in their disutility attitudes towards work and risk). With people's tastes different, it is simply not true that the  $f(D, G, T)$  functions represent well-behaved demand functions with the Slutsky symmetry properties that Harberger invokes and that preference theory needs. This flaw could be circumvented by using the device suggested in my 1955 *Q.J.E.* paper on "Social Indifference Curves." (Let ideal lump-sum redistributions of income take place in the background to ensure there is maximized a Bergson welfare function that requires all men's welfare to rise and fall together along some arbitrarily specified pattern. Of course, this would involve the same partial self-contradiction as does Harberger's case, which tried to tackle the problem of imperfect feasible taxation while assuming in the background that ideal lump-sum taxes were available to keep percentage income distribution constant.) If shares of linear product are kept constant, the S-H integrability conditions will be locally approximated for such an "undercompensated change."

This permits me to end on a constructive note with an exact theory of approximation (no contradiction). Consider an optimal *laissez faire* situation

that maximizes a social welfare function with zero government expenditure and taxes. Now introduce government services as a (vector) function of a small or large parameter  $y$ ,  $G(y)$ . Suppose excise taxes (on goods or services) to be alone feasible, and introduce a (vector) pattern of excises  $T(y)$  sufficient to provide resources for the  $G$ . What is the optimal  $T(y)$  pattern to maximize the welfare function  $W(y) = W(T, G)$ ? This is the problem set by Pigou to Frank Ramsey in the 1920's. Approximate answers were given by Ramsey, Boiteux (and in unpublished form by Hotelling and Hicks). In a pearl cast before the U. S. Treasury in 1950, I gave an exact solution for large and small programs; namely, that the optimal pattern is the one at which the response of all goods and factors to a further compensated-Slutsky price distortion would result in equal percentage (virtual) reductions.

My literary wording is loose, but there is no need for approximative consumer's surplus at all. However, under the stated artificial conditions, some exact statements can be made about small programs (actually about derivatives not finite differences). Expand  $W(y)$  in a Taylor's expansion involving a quadratic plus an exact remainder. If distribution had not been correct to begin with, all consumer surplus terms would be of the second order and negligible in comparison with redistributational effects. (Robin Hood knew that. Does Harberger's reader?) But with ethical optimality realized, the first-power term in  $y$  becomes independent of the tax pattern. The coefficient of the second power of  $y$ , aside from involving one-signed cardinal factors of utility, will be found to involve a factor that looks like a consumer's surplus triangle  $\frac{1}{2}\Delta Q\Delta T$ ; namely,  $\frac{1}{2}\Sigma [dT_i(O)/dy] [dQ_i(O)/dy]$ . This measure of "deadweight loss," which is as exactly defined as Koopmans might want, can be minimized by choice of tax pattern  $[dT_i(O)/dy]$  subject to the constraint that resources are raised for  $G$ . The exact Boiteux solution for this small tax case corrected Ramsey's neglect of income effects; and Harberger's consumer-surplus measures could legitimately arrive at this same result. (He has built well on the upper floor!)

Applying stochastic theory to the "pure theory of public goods," one can derive, by appropriate assumption, conclusions that disagree or agree with Professor Hirshleifer's tentative dictum that government should use the same high interest rates that industry does. The most plausible assumptions do not point his way. I agree with Vickrey that many risks are unavoidable to the private investor and corporation which simply do not exist for We, Inc. The 1923 German inflation wiped out every *rentier*; it left the real economy and taxing power virtually intact. A private investor faces depression risk in every venture. It would be a blessing if the government incurred dollar losses at such times. Only those huge decisions that bring on serious results to all citizens—atomic bomb treaties, and not regional dams—need strong discounting for risk dispersion. (The many projects our large government does may give Koopmans the near-normality needed if Markowitz  $(\mu, \sigma)$  efficiency is to be valid necessary condition. Quadratic utility is an empirically weak reed for the  $(\mu, \sigma)$  theory to rest on. It is a fallacy to think that 2-parameter probability distributions removed from normality provide a valid defense. Even if we forget atomic bombs, and waive cases like the Cauchy



distribution where variances are not finite, being on the Markowitz frontier need not be a necessary condition for utility maximization: an average of two independent variates may have lower expected utility than each separately—even when they each belong to well-behaved 2-parameter families and utility is regularly concave.)

One can look at much of government as primarily a device for mutual reinsurance. General Motors can borrow at a lower rate than American Motors because it is a pooler of more independent risks. It would be absurd for G.M. to apply the same high risk-interest-discount factor to a particular venture that A.M. must apply. The same holds for We, Inc., which is a better pooler of risks than even G.M. (If I thought that bureaucrats made rotten estimates of future benefits, I'd adjust their benefit estimates since it would be accidental in principle if I could achieve the optimal correction by stepping up the time-discount factor. An erratic branch manager of G.M. merits the same readjustment procedure; indeed, raising the interest rate on optimistic fools hampers them little, since they promise you pie in the sky tomorrow!)

If G.M. adds more and more risky investment, that should affect its credit standing. Thus, a risky project may raise the bond yield it must pay from 4.0 percent to 4.0001 percent. The "marginal cost" of the project is not the last low number, but rather that number plus the increase that has to be paid on all the "previous" borrowings. There is also the incremental risk on equity owners due to leverage. The G.M. president will take all such facts into account as will the U.S.A. President.

Obviously, stochastic elements make for bigness and "collectivism." Collectivism need not be governmental; it can also mean monopolies and mutual funds. Often, government is one of the "cheapest" ways of providing insurance against important risks. Why are there no mutual funds that enable me to invest in 15 percent risks (and that could enable society to bring down the cost of risky projects)? If such institutions were efficiently possible and existed, that could drive government out of some activities just as civilian job opportunities drive me out of the peacetime army! If Hirshleifer were right for America with a properly valued dollar, he would become wrong in a situation where there was no feasible way to revalue it. If he were right after a burst of capital-using inventions, he could become wrong in an era where deepening of capital brought down all yields of each degree of riskiness and (by Keynesian reasoning) called for greater collectivism, private and public. The problem is one of costs: What costs in freedom are involved in risk reduction by government? Value judgments and interests are involved. Since unanimity will never be possible, coercion in the sense of acquiescence by dissenters can never be absent; nor can its minimum quantity be defined independently of discordant value judgments.

## EFFICIENCY IN AGRICULTURE

### EFFICIENCY OF THE FARM FIRM<sup>1</sup>

By EARL O. HEADY  
*Iowa State University*

Unfortunately Marshall's typical firm, replicated 3.5 million times, does not exist in U.S. agriculture. Hence, we should ask: which farm firm? We might discuss any of the economic classes of farms defined by USDA: those household firms which decide to sell their labor through other sectors and leave agriculture; the 63 percent of farms with sales of less than \$5,000 in 1959; the 31 percent of family farms with sales over \$10,000 but employing less than 1.5 man years of labor; or the 33 percent of farms with sales over \$10,000 which produce 75 percent of all marketings. The small low-income farms in the South and Southeast or farms operated by nonwhites also represent a more uniform group than "all farms." A large-scale automatized cattle feeding farm in Arizona bears little resemblance to an Alabama sharecropper's firm. Hence, to provide some boundaries for our topic, we refer mainly to the third of all farms which produces three-fourths of sales, but occasionally our reference will be to all farms.

The "statistical average" firm, assisted by a decline in small farms, has been an extremely viable economic unit over recent decades. It has responded greatly to change even though its problems of adjustment have been much greater than for the farming industry. Between the decade 1930-39 and 1960, its increase in land inputs was 4.7 times that of the industry. Comparable figures were 5.3 for farm real estate and 6.0 for all inputs. To make these changes, it also had to increase its investment by a much greater proportion than the industry. Between 1940 and 1960, the average value of assets per farm increased by 469 percent (constant dollars) while assets for the agricultural industry increased by only 284 percent. Output per farm increased by 157 percent over this period, while output for the industry increased by only 56 percent. The magnitudes of changes were possible, of course, only because 40 percent of census farms were no longer represented among agricultural firms as their operators sold their labor elsewhere, retired, or died.

If even more farms had been abandoned as decision-making units,

<sup>1</sup>To be reprinted as Journal Paper No. J-4756 of the Iowa Agricultural and Home Economics Experiment Station, Ames, Iowa. Project No. 1135.

the statistical average farm would have increased its input level further, its scale would have been extended nearer the current least-cost volume, and productivity would have grown at rates even more rapidly than the high rates experienced. These possibilities are apparent: if all agricultural resources had been controlled by farms structured similar to those with marketings of \$10,000 or greater, less than a million could have produced the 1960 output with a work force lower than 4 million. Even then, these fewer and larger units would not have been optimized in the sense of resources allocated to equate marginal value productivities in different investment alternatives and to resource prices.

Slack does exist in the efficiency with which particular strata of farms uses resources, if this slack is measured simply against accomplishments of other farms or in terms of orthodox optimizing conditions.<sup>2</sup> There are restraints and conditions peculiar to individual farms which provide a justifiable basis for "nonmathematical optimizing." But aside from these reasons for an "equilibrium use" of resources departing from static and competitive optimizing conditions, an important question is whether the rate of change in the farm firm should have been any more rapid over the last fifteen years in major farming areas. Recent studies suggest that the typical commercial farm departs a third or more from a scale allowing a fairly full attainment of the cost economies associated with modern machine technology. While the trend is further in this direction, the rate can be speeded only with a more rapid exodus of families and labor from agriculture. Yet it is to be proven that the rate of decline in farm numbers should have been more rapid than has been experienced; or that it should be any more rapid in the decade ahead, especially in extensive farming areas with great spatial separation from concentrated population and industrial areas. The nation's relatively high unemployment rate could be given as one reason. But more important for particular areas, the rate of change probably has been about as rapid as could be digested easily by the rural communities involved. Change in the structure of individual farms has its impact on nonfarm families in rural communities through diminished demand for goods and services, a declining tax base at a time of need for increased investment in education and public services and frequently a general deterioration of services in rural communities. Optimum is a common trade term in economics, but what is an optimum rate of change?

The scale and efficiency of other firms and public institutions are as important for public concern as the farm firm in many rural com-

<sup>2</sup> Even for all U.S. farms in 1959, net labor productivity of Class I farms was three times that of Class II; Class II farms were three times as productive as Class III; etc.

munities. The same problems of scale in relation to economic growth, relative factor prices, and capital technology also relate to the grocery store, machinery distributor, public school, and other economic and social units of the rural community. To serve efficiently in producing services of the standards desired and at unit costs consistent with the technology currently available, these nonfarm firms and service-producing units need to spread out over a much larger volume and geographic area and to make further substitutions of capital for labor. It is likely that the inefficiencies of the nonfarm producing units in rural communities, particularly those provided by the public, are relatively as important as those of the farm firm. Rural communities find it a little easier to assimilate changes in nonfarm producing units and institutions than in accommodating a decline in numbers and expansion in sizes of farms. How much more rapid should these changes be if efficiency is gauged in the broader welfare context which considers the sacrifices and difficulties falling on particular strata of the population in order that slightly (considering unemployment rates) greater increments can be made in the gross national product?

We have discussed some potential aspects of attainable efficiencies in the farm firm. Numerous studies indicate that the earnings of resources in typical farm firms, even for the larger of census commercial farms, are subnormal as compared to other industries.<sup>8</sup> Labor productivity and factor earnings could be pushed nearer to competitive or "equilibrium" levels through use of more capital per farm, and with an extension of size to allow greater attainment of existing cost economics and a lower capital-output ratio. These opportunities will undoubtedly extend for some time into the future. However, there also are other criteria by which efficiency of the U.S. farm firm might be gauged. Whether we decide that its functioning has been "good" or "bad" will depend on the particular criterion applied. Applying the orthodox classroom static equilibrium conditions of competitive firms, we would certainly conclude that slack exists over the extreme majority even of commercial farms. In general, their activities are not opti-

<sup>8</sup>Hathaway (*Government and Agriculture*, Macmillan, 1963, p. 49) indicates that the .8 million farms producing 72 percent of U.S. sales in 1959 evidently had returns on resources below levels of other industries or factor costs. An unpublished study by this writer and Ken Stiles shows Iowa farms with sales of \$10,000 or over failed by 25 percent to cover factor costs with income in 1959. A study by Kaldor and Beneke (*Comparison of Resource Returns on Well-Organized Iowa Farms with Selected Nonfarm Opportunities*, Iowa Agr. Exp. Sta. Res. Bul. 491) showed that even the better organized farms of Iowa did not provide factor returns comparable to other employment alternatives in Iowa in 1954-55. Also see *Farm Costs and Returns, Commercial Farms by Type, Size and Locations* (USDA Agr. Info. Bul. 230, Revised 1963). After covering factor costs other than labor, the latter study generally shows very low labor returns. Imputing market returns to capital for the various economic classes of farms in the 1959 census, we find that only Class I has labor returns equal to or greater than market wage rates.

mized even in the sense of programming models where supplies of some resources are fixed; the majority of programming analysis for typical farms perhaps showing that the net value product could be increased as much as 25 percent, with extremely high returns on capital, against given restraints of land and family labor.

We might also evaluate the farm firm in terms of the level at which it rewards the resources used in it and the extent to which it provides its attached household with consumption items and opportunities consistent with the American standard. Gauged alone in terms of such things as resource awards, level of consumption, and ability to keep family members in school (to gain later access to improved opportunities in economic growth), the farm firm ranks rather low. While the presence or void of certain living facilities and educational attainment or opportunities vary greatly among income strata of farms and locations, rural families show greater school dropouts, fewer modern living facilities, and less satisfactory housing than do urban families.<sup>4</sup> In terms of median money income, farmers and farm managers receive only one-half as much as operatives and a third less than nonfarm laborers. The farm laborer receives only a third as much as the non-farm laborer. However, even though households attach directly to firms in agriculture, these are not appropriate measures for evaluating the farm firm. The economic units and relationships more nearly responsible for the low level of this subset of quantities are to be found in the markets and public institutions which surround agriculture. Depressed resource earnings relate more particularly to the relatively low short-run supply elasticity resources to particular firms in agriculture. The reservation prices of labor, buildings, machinery and agricultural land are extremely low for the farm firms on which they are found, because these resources are so highly specialized towards farming and have low value productivities in alternative employments. Supply elasticities for labor in agriculture, and even indirectly for buildings and machines which would otherwise become abandoned more readily (i.e., have reservation prices falling to zero), can be changed hardly at all by actions of the farm firm; but must be altered through the power of improved alternative employment opportunities, improved and more appropriately directed education, better counseling and guidance services, and increased investment in general adult education for retraining purposes.

### *Other Criteria*

Yet, if our criterion is given an international dimension, with the indices of comparison being the rates of technical advance, growth in

<sup>4</sup>For example, see the data on pages 182-191 in *Agricultural Handbook* No. 258 (USDA, 1963).

supply or its potential, labor productivity or the real farm cost of food to consumers, the American farm firm must be given an extremely high score. Perhaps it even stands at the head of its international class, its only close competitors being farms such as those in Japan and Denmark, which also operate efficiently within their setting of resource prices. Contrasted with Soviet farms of an entirely different scale, management orientation, and organizational structure, and on the basis of the indices mentioned above, the American farm firm must be given a grade of "superior extraordinary." Even in terms of size and scale economics, especially as related to modern machine techniques and labor organization and management, the U.S. commercial farm firm may be nearer the minimum cost optimum than its Soviet counterpart—even though both may depart fairly widely from this mark.

Turning to other criteria of the economics profession, the farm firm must be given a high score in its lack of monopoly profits, if we consider this to be an indication of "goodness." Even under collective action and public legislation to boost prices and income, it has no lasting ability to hold onto returns exceeding the competitive levels imposed by factor supplies which have low short-run elasticities for both the farm and the agricultural industry. Even where competition is removed through quotas and legislated or support prices, interfirm competition simply turns from the realm of products to that of factors which provide the allotment restraints on supply. The intensity of this competition is suggested by the degree of capitalization of price supports and premium from allotments into land values.<sup>5</sup> In this sense, farm entrepreneurs appear to be as rational and "about as quick on the trigger" as investors in securities who quickly capitalize growth prospects and even transitory earning increments into higher common stock prices (although it is unfortunate that these intersector income transfers end up as higher values of land resources rather than for other capital investments, including education and occupational migration). Yet, if the "criterion of achievement" were based on the industrial nonfarm firm's objective of retaining some short-run monopoly or oligopoly (workable competition) profits from innovation, but allowing competition in the long run, the individual firm of agriculture would be considered a retarded country bumpkin by its city cousin.

We mentioned that only a small portion of farm firms are optimized in the sense of resources used to equate their costs with their marginal value productivities. There are important theoretical and practical reasons why this is true, including the interrelationship between the firm and household and the use of an objective function in which profit

<sup>5</sup> Cf. F. H. Maier, J. L. Hedrick, and W. L. Gibson, *The Sale Value of Flue-Cured Tobacco Allotments* (Va. Agr. Exp. Sta. Tech. Bul. No. 148).

does not carry a weight of unity. But even though this is true, the directions of response in supply or production and resource demand or use are remarkably in the postulated directions suggested by changes in relative prices of factors and commodities and marginal productivities or technical coefficients of resources. Substitutions have been made among classes of capital resources and between capital and labor in line with relative changes in their prices and substitution rates. In a similar context, new forms of capital resources have been added, or old forms have been extended, as their prices have declined relative to commodity prices and as their productivities have been discovered to be sufficiently high. Simple review of the mammoth change in structure of agriculture over the last few decades proves this to be true, as does more sophisticated responses studies in commodity supply and resource demand.<sup>6</sup> Adjustment of production and resources to changes in parameters and variables has been strongly in the theoretically postulated and "nationally desired" directions, even if not always by similarly specified magnitudes. (Discovery and insertion of the appropriate lags in the response models would go a long way even in explaining this disparity.) In terms of growth in factor productivity, farm firms have kept apace, or exceeded, nonfarm firms. Figures such as those of Kendrick indicate a considerable margin in favor of agriculture over the two decades 1940-60.<sup>7</sup>

### *Potential Gains from and Needs in Reorganization*

We have mentioned that slack does exist in farm firms when gauged in the usual marginal conditions of hindsight economic equilibrium. How large is this slack and what is the potential gain to consumers in erasing it, when compared to other alternatives in boosting national income? Measurement is extremely difficult and the figures which are available can be challenged. However, it appears that removal of the efficiency gap in farm firms no longer promises to be the nation's eco-

<sup>6</sup> As examples, see the following: Marc Nerlove, *The Dynamics of Supply, Estimation of Farmers' Response to Price* (Johns Hopkins Press, 1954); Earl O. Heady and Luther G. Tweeten, *Resource Demand and Structure of the Agricultural Industry* (Iowa State Univ. Press, 1963); G. E. Shuh, "The Demand and Supply Relation for Hired Labor in Agriculture" (Purdue Univ., mimeo.); Zvi Griliches, "The Aggregate U.S. Farm Supply Function," *Jour. of Farm Econ.*, 42: 282-93; Zvi Griliches, "The Demand For Fertilizer, An Economic Interpretation of Technical Change," *Jour. of Farm Econ.*, 41: 94:103; M. Yeh and Earl O. Heady, "National and Regional Demand Functions for Fertilizer," *Jour. of Farm Econ.*, 41: 332-348; H. W. Halvorson, "Response of Milk Production to Price," *Jour. of Farm Econ.*, 40: 220-231, and G. W. Dean and Earl O. Heady, *Changes in Supply Functions and Supply Elasticities in Hog Production* (Iowa Agric. Exp. Sta. Bul. 471).

<sup>7</sup> See J. W. Kendrick, *Productivity Trends in the U.S.* (Princeton Univ. Press, 1961), pp. 335-65. Extrapolated linearly to 1960, Kendrick's estimates show a growth in productivity: 117 percent for farm, 52 percent for nonfarm, and 57 percent for the private domestic economy. Even using less optimistic figures for agriculture (*Agricultural Statistics*, 1961, Table 662), agriculture compares favorably with Kendrick's figures for the nonfarm economy.

nomic salvation or to save the world from underdevelopment. These potentials were extremely important a century back when the major portion of the nation's capital stock was invested in agriculture and the industry employed nearly two-thirds of the labor force. Now, however, with the labor force and capital of agriculture diminished to around 7 percent of the total, numerous other sectors of the economy pose richer potentials. The unemployed labor force, for example, is one. Others abound. Denison's estimates, which may not adequately consider the impact of existing unemployment or the "fixed" skills of many workers now in agriculture, suggest that the national product might have been a half-percent greater in 1960 with an alternative employment of some resources now in agriculture.<sup>8</sup> Although these figures may be slightly on the low side, they are dismally small relative to alternatives for improvement existing elsewhere in the economy. The urgent problem in resource transfers is not that of overcoming farming inefficiency as a major means of bolstering national economic growth. It is more one of providing greater equality of opportunity for many young persons of farming, so that they are adequately trained to benefit from the employment opportunities stemming from national economic growth in the future. Farm youth and other potential transfers have long been, and continue to be, at a disadvantage in hitching themselves to gains from progress in the nonfarm economy. Aid in decision making is needed more for bringing gain in individual opportunity to these potential transfers than for altering the resource mix of agriculture as a means of major national gain.

As mentioned previously in this paper, response of farm firms to economic stimuli stemming from economic growth have been quite remarkable considering the scale of the farm firm and decisions typically made by a labor manager who is not surrounded by a bevy of economists, operations research personnel, or management specialists. Relative to the management inputs, the extent of decisions "correctly oriented in the direction of change in prices and technical coefficients" perhaps is surpassed in few other segments of the economy. Decisions of the farm firm are not optimum, but neither are those of the Ford Motor Company or similar firms which make large investments in management personnel and sophistication of decision procedures. Within the environment of firm decisions, the efficiency of American agriculture will continue to improve. Improvement will be speeded as operators with least managerial ability and smallest amounts of capital continue to abandon the industry. While we complain of ineffective communication between nonfarm job opportunities and the

<sup>8</sup> E. F. Denison, *The Sources of Economic Growth in the United States and the Alternatives Before Us* (C.E.D., 1962), pp. 198-211.



labor force of agriculture, the movement of people from farms in the next decade will certainly upgrade the managerial level, and the ability of agriculture to expand further and more rapidly in the direction indicated by economic stimuli and technical changes under economic growth. It is not unlikely that the upper quarter of managers then will have greater technical and economic knowledge than the last generation of county agents. In fact, this condition has already been attained by a good-sized strata of farmers in major commercial agricultural areas who no longer depend on the county agent for their supply of innovating knowledge. Cooperative extension services emphasizing agriculture are likely to find the base of their support dwindling rapidly unless they can greatly upgrade the quality of information resources at the disposal of leading farm managers—a trend already somewhat in evidence. State extension services already should begin planning for a time when county offices can be shrunk in number to be manned by persons who are the equivalent of today's extension (or even research) specialist. These local offices will not only need to dispense technology but also must be oriented to provide greater aids in economic and social decisions over both firms and communities.

The declining real price of capital and its substitution for labor and land allows private industry to furnish an increasing proportion of the inputs used by agriculture. Accordingly, the private sector has a more intense profit motive for conducting research to increase the productivity of resources, discover new resource forms, and communicate this knowledge to farm firms. While the private sector plays a small role in research and communication at low stages of development when inputs are largely labor (few inputs can be sold to agriculture), it substitutes effectively for public sector services at high stages of development when capital inputs dominate. This process increasingly takes place in American agriculture and the public sector is freed to shift more of its educational services to economic and social realms of knowledge which cannot be packaged and retailed by the private sector. Progress has not kept abreast of the opportunity in these respects over the last two decades, nor with the relative marginal productivities of public investment in these two realms of knowledge directed at agriculture.

We have mentioned one important public implication of trends which will step up the decision abilities and efficiency of the farm firm. Other implications are obvious. The successful response of farmers to changes in relative prices and productivities of resources will continue the potential and actual substitution of capital for land. The supply potential or surplus capacity of the nation's agriculture will be

maintained or extended and may even intensify the American agricultural dilemma; namely, farmer response to price and technical coefficients in a manner that increases output and resource productivity in the firm but which does not bring greater aggregate revenue to the industry.<sup>9</sup> Even if the number of farms is cut back to a million, with the farm labor force halved again, this potential will still exist. With fewer and larger farms, a smaller aggregate income could represent greater income per farm. Yet the question will remain: how can positive-sum outcomes be guaranteed with (a) consumers realizing continued gains in lower real prices for food and relative declines in the resources needed for agriculture and (b) all farmers realizing some positive benefit from this progress to which they contribute. The intensity of this problem will, of course, depend on the selectivity which occurs as the number of farm firms dwindle towards the million level. Of the firms which continue in agriculture, do the managers which remain, as compared to times of the past, tend to be (a) less conservative in managerial decisions and innovations but more conservative in types of public programs desired or valued, (b) more conservative in farm innovation and less conservative in acceptable programs, or (c) a combination of these two? It is perhaps certain that the very thinning of the ranks of farm managers, as farm firms become more efficient and further extend their scale, leads toward some extreme.

The farm managerial population which is emerging is not a "probability distribution," with individuals changing but the overall frequencies remaining undisturbed. If the process of selectivity results in the first condition, an increased tempo in the technical advance of agriculture can be expected to result in preference for certain formats in agricultural programs. (The selectivity may have less effect in turning farmers completely away from public policies, than in turning their preference to certain types of programs.) We know very little about the managerial selectivity taking place and the impact that it eventually will have on so-called "farm policy values." Yesterday's small-resource, low-income, and somewhat less literate farmer, who looked first to public agricultural policies as a means of welfare improvement, will not be on hand in the future to stump politically for particular kinds of policy. He will either have left farming or be a

<sup>9</sup> Studies suggest that by 1980, in the absence of unexpected demand increments, "surplus capacity" could amount to 50-80 million acres of land which will not be needed for conventional crops such as wheat, feed grains, cotton, etc.; and can be shifted to grass, forestry, recreation or other uses, with the land available for return to crops at other points in time when demand grows sufficiently. For example, see G. A. Selke (chmn.), *A Land and Resource Policy* (USDA); and M. Skold and Earl O. Heady, *Projected Capacity of U.S. Agriculture Considering Interregional Relationships in Production* (Iowa Agric. Exp. Sta. Bul. forthcoming).

part-time operator who looks more to his labor union as a collective means of helping capture his share of the "just fruits" of economic progress.

In recapitulation and elaboration, we believe that the farm firm has functioned reasonably well over the last two decades. Its response has been in the "correct" direction, both to confirm the theoretician's hypotheses in respect to response stimuli and to provide gains to the consuming economy. While its use of factors has not resulted in returns comparable to other sectors and the resources on millions of individual farms could be reorganized to provide a greater product from given inputs (temporarily, for some, a greater value and income), the national society can realize relatively small gains from correcting these imbalances, even though national goals and values may call for their eventual and partial elimination. Agriculture now is too small a portion of the national economy and the rate of national unemployment is too high to allow great gains in economic growth to be realized from stepping up the efficiency and functioning of the farm firm. However, because of individual motivation and selectivity, the functioning of farm firms can and will be improved. The restraints on this improvement hinge less on the decision-making ability and economic rationality of available farm managers and more on factor market circumstances, which affect the rates at which labor migrates from agriculture, thus releasing land to more capable farmers who remain to extend scale of operations, and the supply of capital available for financing expansion and modifications.

As a result of the selectivity process now taking place in types of farm firms and quality of their managers, we see three public problems posed for the future: (1) improvement of factor markets so that structural transformations can take place more readily but with emphasis on providing greater positive opportunity under economic growth for those individuals who have only less favorable (if not unfavorable) prospects in agriculture, (2) restructuring and improving the public research and information services directed to agriculture so that they are more consistent with current and upcoming abilities, interests, and decision-making vigor of farm managers (as well as the resulting problems of farm communities), and (3) devising acceptable policies which do guarantee positive gains from progress of agriculture, both to the selective group of farm managers who will remain and to the consuming and taxpaying public.

## FACTOR MARKET EFFICIENCY FOR AGRICULTURE

By G. S. TOLLEY and B. M. FARMER

*North Carolina State College*

Economic efficiency is a simplifying idea in that concern with it implies asking only one question about a resource situation: whether there are ways to make a greater contribution to total output of the economy.

Diversity and change unsimplify the task of evaluating factor-use for agriculture. Resource adjustments have seldom been made so rapidly as in U.S. agriculture since World War II. Superimposed on the puzzle of understanding recently observed rates of adjustment have been the operation of control programs. Chiefly, these have directly affected land and indirectly affected other productive factors.

Within this setting of large equilibrations in nonfree markets, the present paper considers the usual marginal conditions for efficiency. Quantitative hunches will be given about extent to which the conditions are fulfilled. At the end, comments will be made on government policies.

### *Three Conditions*

Does the remuneration per unit of a factor equal the cost of producing the factor? If factor remuneration approximates value of marginal product (see third condition below), a "yes" answer suggests that the value foregone in bringing an additional amount of the factor into existence may be approximately the same as the value that will be produced by it. In the event of a "no" answer, the difference between remuneration and cost of production is indicative of the amount by which national income could be changed by creation of an extra unit of the factor.

The first condition just considered is relevant to whether there is a tendency to over- or undercreate a factor. The other two conditions concern use of a given total factor supply.

The second or factor fluidity condition will be of most central concern: Is remuneration of a factor the same in all its alternative uses? In conjunction with the third condition, differences in factor remuneration indicate gains in product from reallocations of the existing supply of a factor.

Are amounts of a factor, devoted to a particular use, adjusted to the point where remuneration of the factor equals value of marginal

product? This question is relevant to the third or use-value condition. The rationale is that, if the factor remuneration accurately reflects value of goods or services foregone (as implied by conditions one and two), the difference between remuneration and value of marginal product indicates addition to national income from changing the amount of the factor devoted to the use in question. Reasons why market behavior could lead to lack of fulfillment of the third condition will be brought out.<sup>1</sup>

### *Purchased Inputs*

Rates of return of firms in the nonagricultural sector supplying farms with purchased inputs do not appear out of line with those for all manufacturing corporations.<sup>2</sup> This suggests that the factor creation condition, pertaining to factor price in relation to cost of production, may be reasonably well met for these inputs. Although there is sometimes the appearance of circumstances conducive to imperfect competition, such as few producers or sellers, temporary emergences of excess capacity and lively price competition that can verge temporarily on chaotic are indicative of lack of effective collusion.<sup>3</sup> These events mainly suggest growing pains to be expected of competitive industries which in the past fifteen years have supplied a 50 percent increase in farm inputs of mechanical power and machinery and a 120 percent increase in fertilizer and liming materials accompanied by down movement in the real cost of fertilizer.<sup>4</sup>

The lack of suggestions that there are persistent areal price differentials due to monopoly-like discrimination may be indicative that the fluidity condition, pertaining to allocation of a given supply among alternative uses, tends to be fulfilled.

Consider the demand for the purchased inputs, having to do with

<sup>1</sup> Within-firm inefficiencies could also lead to lack of fulfillment of the third criterion. These are not of primary concern in this paper. It should be noted that only if firms tend to make factor remunerations equal to value of marginal product will the differences, involving relation of remunerations to estimated effects on national income, be relevant in considering the other two criteria. The condition that factor remuneration approximates value of marginal product is not as damaging to remuneration comparisons as might seem at first, if one reflects that most within-firm inefficiencies may have to do with reaching the highest production frontier rather than with finding the optimum factor use given the firm's existing production frontier. The third criterion pertains to the actual change in product that would result from a unit change in a factor; it does not pertain to the hypothetical highest potential marginal product to be obtained if firms were to adopt different production practices.

<sup>2</sup> See Federal Trade Commission, *Profit Rates of Manufacturing Corporations 1947-62* (Washington: Government Printing Office, 1963), pp. 22-37.

<sup>3</sup> A recent airing of issues is contained in the paper by Robert F. Lanzillotti and the discussions of this paper by S. Kent Christensen, George R. Slater, and Elliott S. Clifton at the session on market power in the proceedings issue of the *Jour. of Farm Econ.*, Dec., 1960, pp. 1228-1264.

<sup>4</sup> Economic Research Service, *Changes in Farm Production and Efficiency, 1963*, USDA Statistical Bulletin No. 233 (Washington: Government Printing Office, 1963), p. 47.

the use-value condition (i.e., whether purchased inputs are adjusted by farmers to the point where cost equals the value of marginal product). A consideration tending to make for too little use of purchased inputs is high interest rates on agricultural loans. A casual look suggests that these are higher than justified by purely risk considerations; there appear to be inefficiencies connected with institutional arrangements in the financial industry. However, any such inefficiencies are less important than they used to be, as a result of government-sponsored land banks, the Farm Credit Administration, Farmers Home Administration, and other lending agencies which have sometimes overcompensated with quite low interest rates. Often these agencies, whatever their original purposes, have with the passage of time come to behave very much like private financial institutions and today are probably helping to ensure that competitive rates of interest are available to farmers.<sup>6</sup>

Internal and external capital rationing have received attention as considerations in addition to interest rates that might make marginal product of inputs greater than their price. Subject to empirical verification, it may be hypothesized that there is nowadays little irrational capital rationing. That is, most farmers with apparently small amounts of capital would not find themselves able to use more capital to obtain product as great as price of inputs. They might bankrupt themselves. Bankers and the farmers themselves sense it. This correct sensing of low managerial ability can be mistaken by the outside viewer as capital rationing, if the viewer assumes an unrealistic level of managerial ability.<sup>6</sup>

These considerations relating to the third condition for purchased inputs are in the realm of oft-voiced beliefs. A subjective crude guess is that each of them reflects some malallocation but that elimination of them would lead to at most a 2 or 3 percent gain in product obtained from using purchased inputs.

### *The Human Factor and Land*

The rest of the agricultural inputs will be considered together because a burden of this paper is that the efficiency with which they are used depends on joint supply considerations, as will be brought out when considering fluidity and use-value conditions.

*Factor Creation.* The marginal "costs of producing" human inputs

<sup>6</sup> These comments are generally in line with conclusions reached by the authors on agricultural agencies in *Federal Credit Agencies*, prepared for the Commission on Money and Credit (Prentice-Hall, 1963), pp. 259-434.

<sup>6</sup> See *Capital and Credit Needs in a Changing Agriculture*, E. L. Baum, H. G. Diesslin and E. O. Heady, eds. (Iowa State Univ. Press, 1961), for examples of the variety of opinions and research analyses that exist for farm use of credit.

for agriculture are the educational investments that train people to work in agriculture. Particularly in the South, there tends to be relative overinvestment in training the rural youth who are potential farmers in agricultural pursuits while at the same time there is underinvestment in total education for those young persons.<sup>7</sup> The handwriting could hardly be clearer that farming more and more will return a living comparable to that which could be earned in nonfarming only for those with decision-making abilities to manage tens and hundreds of thousands of dollars of capital. Training in a few farm tasks can hardly supply this. Farm management studies show the large farm to be the farm of the future. The doubling of farms with sales of \$10,000 and over between 1950 and 1960 while the total number of farms was declining nearly a third is corroborative.

A conservative estimate from recent studies is that less than one in three rural farm youths will remain in the farm population after they enter the labor force.<sup>8</sup> The expected migration is especially great for nonwhites. These findings have implications for the kinds of educational efforts needed for efficient utilization of human resources. The relative lack of education of rural Negro youths is already being reflected in high unemployment rates for young Negroes migrating to cities. Here is truly a problem of economic efficiency whose importance for the nation as a whole is increased by its widespread geographical effects due to migration.

The supply of land is influenced by government-sponsored flood control for agricultural land, increasing its average productivity by reclamation and by terracing and soil-conserving plant rotations. Private leveling, clearing, and draining are also significant. It has been correctly pointed out many times that direct costs typically substantially exceed the direct benefits of most government land creation activities; i.e., first criterion is not fulfilled in that cost of production of factor exceeds marginal product. At the same time, the well-documented existence of millions of acres of potential agricultural land now in trees suggests that the long-run supply of land on the part of private decision-makers is virtually perfectly elastic. Over a ten- or twenty-year period the effect of government land development is mainly to result in less investments to keep up existing land or develop new

<sup>7</sup> This subject would in a broader setting be considered as part of the general problem of education and efficiency. See T. W. Schultz, "Education and Economic Goals," *Educational Needs for Economic Development of the South*, Agricultural Policy Institute and Southern Regional Education Board, published at Raleigh, 1962, pp. 5-12.

<sup>8</sup> C. E. Bishop and G. S. Tolley, "Manpower in Farming and Related Occupations," *Education for a Changing World of Work, Appendix II, Report of the Panel of Consultants on Vocational Education requested by the President of the United States* (Washington: Government Printing Office, 1963), and W. E. Johnston, "The Supply of Farm Operators" (Ph.D. thesis at North Carolina State, 1963).

land by private decision-makers. Privately developed land will expand, less rapidly than it would have otherwise; or if the total amount of land is declining, some land will go out of cultivation when it would not otherwise have. These assertions assume, first, that the demand for farm products is too inelastic for government-sponsored land development to result in any appreciable increase in total sales of agricultural products—so that increases of production on government-sponsored land must eventually result in decreases of production from other land. Second, flood-control investments and reclamation are costly investments of a fixed nature such that once made the agricultural land is valuable relative to other agricultural land even though costs of development exceed its value. Therefore, lower rent land other than government-developed lands will go out of production.<sup>9</sup>

The usual type of benefit-cost analysis, when direct costs and direct benefits are used, approximate the benefits of land development as the increment to land returns and approximate the costs as market costs of construction and related activities required to develop the land. Direct net benefits as usually measured thus appear to estimate national income effects of government land development for the long-run period. The procedure is consistent with the comment made above that the effect on national income of nonfulfillment of the factor creation condition is reflected in differences between factor return and cost of production of factor. In the short run, the benefits are probably less than suggested by land return increments on the government-developed land because there is a tendency not only to drive low-rent land out of production in other parts of the country but also to displace the farmers on such land. They will not move instantaneously into nonagricultural production, so that there will be tendencies toward surplus production with low marginal value and involuntary unemployment of farm people due to lack of markets.

*Fluidity Between Agriculture and the Rest of the Economy.* With regard to human inputs, it is helpful to think of three poles which in reality shade into one another. First, there are low-income farmers. They produce little and compete for few land or other agricultural resources. Nonfarm events may affect their decisions to stay or not to stay in agriculture, but they have too little commercial connection with agriculture to be much affected by it. A second group is composed of commercial farmers who are at least relatively inefficient and have neither an absolute nor a comparative advantage in agricultural production. They differ from the low-income farmers in that they compete significantly for land and other agricultural resources and are

<sup>9</sup> For a further analysis, see G. S. Tolley, "Interrelated Land Development Possibilities," *Modern Land Policy*, ed. Halcrow (Univ. of Illinois Press, 1960), pp. 121-42.



affected by and themselves affect significantly agricultural supplies, prices, and incomes. The third group is composed of relatively efficient farmers who have both an absolute and a comparative advantage in agriculture. Agriculture is apparently tending toward an equilibrium where the first two groups will eventually move entirely into non-farming, and farm operators are composed entirely of the third group, who will significantly expand in number.

The fact that the equilibrium is not achieved instantaneously is suggestive that the fluidity condition is not fulfilled; that is, it appears comparable labor does not tend to receive equal returns as between farming and nonfarming. Reliable sufficiently microincome comparisons are still lacking, but the available income data are suggestive of income disparities for the three polar groups. Thus the majority of farmers receiving incomes below that in nonfarming represent low-income and inefficient commercial operators, while those prospering are apparently the emergent efficient operators increasing in number. Johnson found that farm and nonfarm people appear to have comparable earning capacities, so that income differences should not be attributed to differences in abilities.<sup>10</sup> Loss in national income due to the lack of instantaneous adjustment can, and later in this paper will, be estimated.

To know what policy view to take toward the apparent nonfulfillment of the fluidity condition for the human input requires an understanding of what determines its magnitude through time. The immobilities preventing instantaneous adjustments do not appear to be importantly associated with costs of migrating nor with missing knowledge of higher paying nonfarm alternatives. Since 1940, a question in the Census of Population has pertained to where each person was at a previous date (either one year or five years in the past). From the answers, flows of people between areas of the country can be calculated. The gross flows, i.e., A to B and B to A, have been found to be several times as large as the net flows. In short, people are moving around the country back and forth at a great rate and not just making once and for all moves as would be suggested if costs of traveling were a main deterrent to moving or if lack of knowledge of opportunities were deterring migration. A regression analysis<sup>11</sup> indicates that distance deters gross migration between states to some extent. However, the most important variable explaining gross migration between a

<sup>10</sup> See D. Gale Johnson, "Comparability of Labor Capacities of Farm and Nonfarm Labor," *A.E.R.*, June, 1953, pp. 296-313, and "Policies to Improve the Labor Transfer Process," *A.E.R.*, May, 1960, pp. 403-12. A review of research on income comparisons is contained in the paper by G. Edwin Shuh, "The Agricultural Input Markets—A Neglected Area of Agricultural Policy and Economic Research," presented at meetings of the Western Farm Economics Association, 1963.

<sup>11</sup> Being carried on by Gordon S. Sanford.

given state and other states is the number of persons born in that state and currently residing in each of the other states.

A study of the supply of farm operators helps to verify in quantitative terms the role of age as a key determinant of mobility.<sup>12</sup> Life career experiences are reflected in the pattern obtained by tracing through time the number of farm operators born in a given decade. The usual pattern has been that the number of operators born in a given decade increases with the passage of time up until these persons are about forty years old. Then the number declines, and progressively so, as they age further, due to retirement and death. The relative stability in cohort pattern reflects the fact that a chief adjustment to the decline in total number of farms has been decreased entry rates for young farmers. When there is a decline in farm numbers, the groups who are already embarked on a career cycle have a tendency to continue the pattern already started. The decline is reflected most importantly in a reduced cohort size for the group just beginning farming.

Most of the variation in cohort patterns can be explained by regression as responses (*a*) to decadal changes affecting total farm numbers and (*b*) to changing ages of rural farm males who provide potential supply of operators. While the major effect of a change in farm numbers is to influence the number of operators under twenty-five years old, the number from twenty-five to thirty-four is also quite sensitive. Then the response declines sharply with increasing age.<sup>13</sup>

The age selectivity and the high response of youth indicate that young people entering the labor force go primarily where they can get jobs. This tends to give the lie to the idea that income differences are due to taste. At the extreme, relatively low incomes in farming might be attributed to leisure and nonmarket choices. Then income differences would simply be the amount that people are paying for outputs that are really a part of national income but do not get counted in the usual measurement of national income. If, as appears more likely, the income differences that persist for people of middle age are due to fear of pain of change, then policies that ease change can be called for as national income increasing. So also can policies that would help prevent career choices later regretted due to lower earnings than expected.<sup>14</sup>

Consideration of reversibility of taste changes further strengthens the arguments. Suppose people did not choose to move in response to

<sup>12</sup> See Warren E. Johnston, *op. cit.*

<sup>13</sup> See D. Kanel, "Farm Adjustments by Age Groups, North Central States," *Jour. of Farm Econ.*, Feb., 1963, pp. 47-60, and M. Clawson, "Aging Farmers and Agricultural Policy," *Jour. of Farm Econ.*, Feb., 1963, pp. 13-30, for similar findings about farm operator age cohorts.

<sup>14</sup> A rationale for policies based on regretted career choices is presented in T. W. Schultz, "A Policy to Redistribute Losses from Economic Progress," *Jour. of Farm Econ.*, Aug., 1961, pp. 554-65.

existing income differentials but that if somehow they happened to move, they would be glad. Furthermore, suppose that if they were in the nonfarm sector and moved to the lower-income situation in the farm sector, they would be sorry. Then in spite of the change in tastes in the two situations one could say unambiguously that they would be better off by moving to the nonfarm sector.

The first of the three polar groups mentioned previously—low-income farmers—are responsive to off-farm opportunities, as shown by several studies which find off-farm migration to rise in booms and fall in busts.<sup>15</sup> Combining this finding with the findings on age, a conclusion is that there must be a ratchet effect of recessions on entering farmers. During recessions, youths leak into low-income farming because of lack of other opportunities but then are not prone to get out once other opportunities emerge later. The second group—relatively inefficient commercial farmers—have similar mobility characteristics, but there are more ramifying consequences within agriculture because they produce significant quantities and are likely to be farming land which the third group—relatively efficient commercial operators—wishes to have in order to expand. So the efficient and inefficient operators are in competition for land and allotments. The concentration of relatively young operators among farms of higher economic class is evidence of the existence of the third, in-moving group.

Many new production practices are being adopted by a majority of commercial operators both efficient and inefficient. This is further heightening competition. Consider the across-the-board additions to production due to adoption of new production practices by virtually all commercial operators plus the additions to production from entering and expanding high-management-level farmers. If the supply of high-management-level farmers were not too large, the foregoing additions to production could be less than production withdrawals due to normal exit rates associated with death and retirement for those already in farming. Then, in a full employment economy, adjustment could be accomplished relatively easily (i.e., with relatively few farm-nonfarm income disparities) through reduction of youth entry into farming. We do not have such a fortunate concatenation. Across-the-board adoptions of new techniques plus a sufficient supply of entering high-management-level operators plus a nonebulient nonfarm labor market have combined to bring pressure, not only to reduce youth entry, but to force out some who have already embarked on farming careers. The latter are vigorous in their competition for land. The rate of adjustment to the new-type agriculture of highly efficient operators

<sup>15</sup> A review and recent contribution is contained in E. O. Heady and Luther G. Tweeten, *Resource Demand and Structure of the Agricultural Industry* (Iowa State Univ. Press, 1963), Chaps. 8 and 9.

depends on the differential product efficient operators can obtain on land and hence bid for it, as opposed to the income sacrifice that operators tending to be forced out are willing to take.

In contrast to the human input, the fluidity condition for land as between agriculture and nonagriculture may tend to be fulfilled. Urban demand for land is inelastic even at very high land prices. It therefore tends to pre-empt agricultural uses. Agricultural demand for land becomes inelastic at lower values than does urban demand but at values several-fold higher than those at which forest demand becomes elastic. Agricultural demand therefore pre-empts forest demand. These demand characteristics give strong incentives to move land into pre-empting uses promptly.

*Fluidity within Agriculture.* With regard to human input, it is conceivable that the South has some climatic advantages for livestock production such that the able cattle and hog producers of other parts of the country could make higher net incomes if they located in the South. However, they may be too immobile to do so. There may be some instances of failure to match up specialized natural locational advantages with specialized managerial abilities due to human immobilities, but a hypothesis is that their efficiency effects are small due to the many alternative places in the country that the major products can be produced at fairly similar cost and due to belief that farmers will migrate around the country to other farming opportunities if income differences are substantial enough.

Hired migratory labor is used impressively efficiently in terms of the careful treks of workers following harvest dates. The side effects in terms of social conditions are another matter not evaluated here.

The fluidity condition for land as it pertains to alternative uses within agriculture is kept from fulfillment partly by government programs. Due to control programs, production has not moved—or has not moved as promptly—between areas in response to changing cost conditions as efficient use of land in terms of the second criterion would require. But the quantitative effects on total income may be limited due to the general similarity of production costs between regions even where there are incentives to shift.<sup>16</sup> The detractions from total product due to factor substitutions induced by land price effects of programs appear even smaller.<sup>17</sup>

Consideration of land allocation within agriculture returns us to the

<sup>16</sup> By comparing total costs of producing agricultural outputs under alternative geographical distributions of production, estimates could be made of the quantitative effects on total income of retarding geographical shifts. The definite indications that programs affect the geographical distribution of production suggest the worthwhileness of doing so.

<sup>17</sup> There have been studies substantiating that these factor substitutions add little to the cost of producing agricultural products. See the review at the Pittsburgh session on "Principles of Economic Policy, Consistent and Inconsistent: Economics of our Present Farm Price Support Policy," printed in the *Jour. of Farm Econ.*, May, 1963, pp. 331-58.

discussion of the force-out. A part of the group of relatively inefficient commercial operators who are tending to be forced out owns the land they are farming. If they had to go into debt to acquire land to farm or if they had to compete with group three in renting, many would not do so in view of the low net income that would then be available to them for family living. This is the same as saying that adjustment is slowed by high reservation prices for land on the part of present owners. The reluctance to sell is reinforced by past acquisitions of capital fixed assets.<sup>18</sup>

Capital gains resulting from competition for land do, however, help overcome immobility. The only way a person can realize these capital gains is to get out of farming. (If he acquires another farm, he will have to pay the same high land prices he sold for, leaving his balance-sheet situation about the same.) As land values rise, the income that could be obtained by investing the proceeds of the capital gain are an increasing incentive added to any income from nonfarming pursuit that would be earned from quitting farming. One of the capital gain effects may be to induce earlier retirement.

*Use-Value.* Factor remuneration may not always equal value of marginal product for human inputs and land due to rigidity of share tenure arrangements. These can have suspicious fixity over long periods. Their fixity may be a friction. In a hypothetical perfect market, as factor demands and supplies shifted, factor shares would change. With rigid factor shares, the shifts lead to such phenomena as surpluses of persons seeking farming opportunities at going shares or an excess of farms for sale or rent. But adjustment in side arrangements will tend more and more to be made the farther the drift from equilibrium. There will be more cash arrangements. Factor shares have changed accompanying significant changes in technology and farming patterns in an area. At such time when there is a total reconsideration, shares tend to be set up close to the then equilibrium from which they start drifting again. The drifts will tend to last longer in tradition-oriented areas. The probable lack of much substitutability between factors in the range of drift means that factor-use is probably not very much affected. Thus, the departure from equilibrium appears more a matter of division of income than an important effect on resource use.

### *Savings from Equalizing Earnings*

The discussion has suggested that the major nonfulfillment of conditions for efficient factor use is the lack of comparable earnings of human inputs as between farming and nonfarming. An attempt has been

<sup>18</sup> See Glenn L. Johnson, "The State of Agricultural Supply Analysis," *Jour. of Farm Econ.*, May, 1960, pp. 435-52.

made to contribute to understanding of the rate at which earnings tend to equalize—stressing age, management abilities, and joint supply characteristics of the human and land factors. It does not appear possible as of this writing to accurately predict the rate of adjustment.

Given the hypothesis that human resources with higher management abilities have comparative advantage in agriculture, an estimate can be made of the resource savings that would result if it were possible to bring the human factor into complete adjustment. Suppose the end result of the adjustment was that the remaining farmers all had man-hour efficiency equal to that being achieved by the better farmers of today. The following table shows savings in man-hours that could have been achieved if outputs had been produced with labor coefficients from farm budgets for better farms in comparison to the actual man-hours used in 1959 as reported by USDA.<sup>19</sup>

	Millions of Man-hours Used in 1959	Percent Savings If Labor Had Been Adjusted According to Static Criteria
Feed grains.....	840	35-60%
Hay and forage....	517	10-25
Food grains.....	190	40-55
Cotton.....	911	50-80
Tobacco.....	539	20-40
Oil crops.....	161	35-50
Meat animals.....	1,395	20-45
Dairy products.....	1,841	15-30
Poultry.....	657	55-70
Other crops.....	1,199	—
Overhead labor.....	2,051	—

The total savings for the lower range of estimates is 2,050 million man-hours and for the upper range is 3,400 million. This compares with total man-hours used in 1959 of 10,300 million. The savings would be greater if estimates were available for the many minor other

<sup>19</sup> The source of the 1959 man-hours is *Farm Production and Efficiency*, op. cit., p. 36. To estimate the percent savings in labor, budget studies for major areas of production of commodities were obtained. Over fifty studies, published by Departments of Agricultural Economics, were consulted. It was assumed that labor coefficients (man-hours per unit of output) from these studies reflect the efficiency of higher levels of management. Where budget studies were not available for all major production areas, nearby areas were selected as similar as possible in agricultural characteristics. Hypothetical man-hours were obtained by multiplying the labor coefficients by output of each commodity as reported in the USDA publication *Agricultural Statistics*. The labor savings are the differences between the actual 1959 man-hours and the hypothetical man-hours.

The ranges result from variation among studies in man-hour coefficients for the same or similar areas, and they also reflect allowance that was made for different geographical distributions of production. There was no attempt to estimate the most efficient geographical pattern of production, but it was recognized that alternative geographical patterns affect total man-hours to some extent. For instance, the range for cotton reflects alternative assumptions about the proportions of production taking place in the South and the West.

crops and overhead labor. Estimates for these would require a greatly extended analysis. As an order of magnitude, then, if the saved labor could produce between \$1.00 and \$2.00 per hour in nonfarming, the savings for that year would have been at least 2-7 billion dollars.

### *Policy Addendum*

Policies to encourage more rapid adjustments include general monetary-fiscal measures and local area development activities to increase availability of nonfarm employment. They can be recommended as a help. The remarks on rate of adjustment make clear that increased off-farm availability of jobs will not alone accomplish anything near instantaneous adjustment. The most pointed programs would be vocational rehabilitation programs for farmers without a future who have already entered. The larger the vocational measures the better, at the same time realizing there is no realistic hope that they will achieve complete adjustment. Virtually all measures to speed labor adjustment can be recommended except indiscriminate lowering of prices which through capital losses and forcing to the wall income-wise might be effective.<sup>20</sup>

Lags in factor adjustment, accompanied by agricultural surplus tendencies and income differentials, appear to be in prospect for at least several more years. We need dynamic criteria of efficiency as a guide to how fast to expect adjustment. To ask only to go toward fulfillment of static criteria is too weak. To ask for instantaneous adjust-

---

The percentage range estimates have all been rounded to the nearest multiple of five.

The estimated savings pertain only to how to produce a given output more efficiently, not to savings that could be realized from a more optimum amount of production of agricultural goods. The latter is a matter of product-market not factor-market efficiency; but one may note that, due to price inelasticity of demand, production cannot get very far from optimum if there is not surplus accumulation and decumulation. The above estimate has the advantage of avoiding ambiguity about how to value marginal agricultural output, since it considers gains in product holding agricultural output constant. This is an important advantage in view of the ambiguity as to marginal value where there are surplus tendencies and prices are influenced by government programs.

The fact that with high-level management more purchased inputs would be used after the labor adjustments means that considering labor alone slightly overestimates the gains in total product from adjustment; i.e., while labor is released from agriculture to the rest of the economy, more purchased inputs are taken out of the rest of the economy. This does not appear to be a serious limitation because farm inputs from the nonfarm sector are only about one-sixth of total farm inputs; their increase associated with the labor adjustments would be only a fraction of this amount.

The labor adjustments would probably lead to some increase in total hours worked, inasmuch as people transferred to nonfarming tend to work more hours than on farms where there is seasonal idleness. The gain in money national income from any increase in total hours worked is not taken account of in the text estimate. The money gain would be a real gain if, as seems reasonable, a near-zero value is placed on the leisure given up.

<sup>20</sup> Lower prices would tend to drive out efficient as well as inefficient operators. The effect of product prices on rate of adjustment depends on programs accompanying the price change, most particularly on whether there is still production control affecting nature of competition for the land factor. It is not clear that lower prices would be effective.

ment is asking for too much. The nearly 50 percent decline in labor inputs in agriculture in the last fifteen years is evidence of substantial adjustment somewhere between the too-weak and too-strong rate.

We need even more a framework to help in (*a*) balancing dynamic efficiency criteria along with other goals and (*b*) considering goals in the context of operational policy-making procedures. Lacking this framework, can we not say intuitively that a satisfactory degree of efficiency has been achieved and is likely to be in the future under a wide range of policy alternatives?



## DISCUSSION

K. L. ROBINSON: While very substantial gains in efficiency have been achieved by farmers in the United States during the past two decades, the evidence presented by both Mr. Heady and Mr. Tolley indicates that the adjustment process is far from complete. With a more nearly optimum number and size distribution of farms than at present, and no further gains in technology, labor requirements in agriculture, according to their estimates, could be reduced by a minimum of 20 percent and perhaps by as much as one-third. Although American agriculture is clearly not as efficient as it might be, both authors raise questions as to whether society would have gained significantly by increasing the rate of movement out of agriculture above the very high levels that have prevailed in recent years. Few of us, I think, would be prepared to argue that the gross national product would have been significantly higher in the past decade if more resources had moved out of agriculture. And certainly the social and economic problems of rural communities would have been even more acute if the rate of migration had been accelerated. Unless unfilled demands exist among nonfarm firms for labor and land, there is clearly little to be gained by releasing additional quantities of these resources from agriculture.

Undoubtedly there will continue to be a high rate of movement out of agriculture in the years immediately ahead, contributing to what Mr. Heady calls "managerial selectivity." The policy issues raised by selective migration out of agriculture as well as selective entry into farming are indeed interesting ones. Mr. Heady performs a useful service in calling attention to the impact this may have on the Extension Service and on the demand for the services of agricultural economists as well. He also suggests that managerial selectivity may ultimately affect farm policies. Based on limited observations of what might be called the new "managerial class" of farmers in the Northeast, I would expect the majority of these to oppose compulsory supply control programs. The new managers are generally "expansion-minded" and somewhat less concerned about moderate price reductions than many of their neighbors.

Mr. Tolley suggests that capital rationing is no longer a major problem in American agriculture. If so, much of the literature in agricultural economics needs to be revised. I welcome this statement, especially coming as it does from a distinguished graduate of the institution where the concept of capital rationing in agriculture has received so much attention. My own view is that significant changes have occurred during the past two decades, both in the attitudes of creditors towards lending to farmers and in the attitude of farmers towards borrowing. In conversations with bankers, managers of cooperative credit agencies, and even farmers, I have gained the impression that creditors may even err on the side of being too liberal in granting loans at present. Competition among banks and other credit institutions for additional business

had led to making loans that earlier generations of creditors would have considered "unsound."

One also may raise questions about the marginal value to society of increasing the total amount of capital invested in agriculture, particularly if the additional capital simply leads to higher land values or to additional farm output (which may result in additional public costs under present support programs). The interests of society clearly may conflict with those of individual farmers. To increase efficiency and lower costs, many farmers need to make additional capital investments. If these investments result in a substitution of capital for labor without any increase in output, or if output increases and prices reflect, in part at least, the lower costs of production achieved by increasing efficiency (or the labor released from agriculture contributes to the total product elsewhere), society as well as the individual farmer making the investment obviously will gain. But if the released resources remain unemployed and output increases, but because of support programs, society fails to realize the potential gains from lower production costs or is forced to pay for diverting more land or commodities to secondary uses, marginal returns from increasing the aggregate amount of capital invested in agriculture may be zero or negative.

Mr. Tolley likewise raises some questions about the returns to additional public investments in land. As with returns from other forms of capital invested in agriculture, conflicts may rise between the profitability of such additional investments to individuals, groups or areas, and the returns to society as a whole. If additional output is valued at average prices, then reclamation schemes or subsidies for lime and fertilizer appear to be profitable. But if these public investments result in additional output and such increments are valued at their marginal returns to society, then few of these public investments in agriculture would be justified. Society may have to pay, not only the cost of the subsidies for construction, but also the marginal costs of taking more land out of production, storing and diverting additional quantities of supported commodities or restricting imports, as of sugar for example. In a technical sense, many of these public investments lead to higher efficiency, that is, more output per man, but we should not be misled by average figures. We must examine the more indirect consequences, as Mr. Tolley emphasizes, if we are to allocate public resources in the most efficient manner.

Mr. Tolley tends to minimize the possible adverse effect of acreage allotment programs on agricultural efficiency. He concedes that gains in efficiency might be achieved by shifting production between areas, a process that has, for some crops, been retarded by allotment programs; however, he says the "quantitative effects on total income are probably small due to the general similarity of production costs between regions." Surely differences in production costs of as much as 20 percent are not insignificant. In the case of cotton, Mr. Tolley suggests potential savings in labor inputs of as much as 50 to 80 percent. Only a small part of these potential savings could be achieved without interregional shifts in cotton production, as he himself recognizes.

Very little was said in either paper about the possible influence of govern-

ment price-support programs on efficiency in agriculture. This is a much debated subject, but the available evidence suggests that total inputs of both labor and land would not have been much different during the past decade if price supports had been eliminated. Shifts between crops obviously would have occurred in the absence of supports, but the aggregate quantities of these resources committed to agriculture probably would have remained about the same due to the highly inelastic short-run supply schedules for both land and labor presently employed in agriculture. The major impact of lower prices would have been to reduce the returns to these resources rather than the quantities employed. Quantitative effects on inputs probably would have been concentrated mainly on those items which are purchased from nonfarm industries, including irrigation equipment, building supplies, and machinery.

It is probably fair to conclude that nonfarm policies, particularly those designed to increase employment opportunities and improve the education and skills of those living in rural areas, will have more to do with future gains in agricultural efficiency than will so-called "farm programs." The latter will continue to influence the quantities of resources purchased from nonfarm industries and the returns to labor and land already committed to agriculture but are likely to have only a modest effect on farm efficiency.

# EFFICIENCY IN THE LABOR MARKETS

## RELATIVE EMPLOYMENT EFFECTS OF UNIONISM

By H. G. LEWIS  
*University of Chicago*

1. Although the effects of labor unions on wages and employment in the U.S. have been a subject of speculation among economists for many years, our stock of estimates of the numerical magnitudes of these effects is mainly a product of research reported in the last fifteen years. In *Unionism and Relative Wages in the United States*<sup>1</sup> I have taken an inventory of the empirical literature, comprising about twenty studies, on the relative wage effects of unionism in the U.S. I have estimated from the evidence in these studies that the impact of unionism on the average wage of all union labor relative to the average wage of all non-union labor may have exceeded 25 percent near the bottom of the Great Depression of the 1930's, was 5 percent or less in the latter 1940's, and about 10 to 15 percent a decade later.

Twelve of the studies provide estimates of the effects of unionism on average relative wages for eighteen different industries or significant occupational groups within industries,<sup>2</sup> although not all of this detail is available at any single date. The range of these estimates for the period after 1935 is from near zero to close to 50 percent. However, most of the estimates are below 25 percent, and I would put the dispersion (as measured by the employment weighted standard deviation) of the effects of unionism on average relative wages among all industries in the latter 1950's at 4 to 6 percent. I regard this estimate of the interindustrial dispersion as highly tentative and have not attempted to make even such global estimates for the distributions of the relative wage effects by occupation, locality, and other ways of dividing the labor force.

2. The production of the present stock of studies of unionism and wages followed fairly closely after and surely was stimulated by the great increase in the fraction of the labor force represented by unions that occurred in the decade 1935-45. However, the growth of unionism

<sup>1</sup> Univ. of Chicago Press (1963).

<sup>2</sup> Wage earners in bituminous coal mining, men's clothing manufacturing, steel manufacturing, and rubber tire manufacturing; employees in the manufacturing of paints and varnishes, footwear, cotton textiles, automotive parts, wooden furniture, hosiery, and women's dresses; hotel employees; barbers; physicians; commercial airline pilots; local transit motormen; seamen in East Coast shipping; and skilled and unskilled building trades.

in the economy did not also stimulate a substantial postwar output of numerical estimates of the relative employment effects of unionism.

I do not mean to suggest that there has been no empirical work dealing with the impact of unionism on relative employment. Indeed, it would not be difficult to assemble a rather long list of works that inquired into one or another of the employment aspects of unionism.<sup>3</sup> With rare exceptions, however, one cannot transcribe directly from these studies, or with little effort produce from the data in them, numerical estimates of the size of the employment effects. The exceptions, in my judgment, provide too small a base for global numerical statements about the impact of unions on the allocation of labor in the economy.

3. In the remaining part of this paper I present some new numerical estimates of relative employment effects of unionism among industries. One set of estimates—that discussed in this section—pertains to a very broad division of the economy into industries. The second set, not yet completed, is for bituminous coal mining, for which I suspect the employment effects are among the largest in the economy.

I have divided the civilian (nonwork-relief) economy into two industry groups, Group *a* containing the mining, contract construction, manufacturing, transportation, communications and public utility industries, and Group *b* all other industries. In the period 1919–58 more than 80 percent of the union members in the economy were employed in Group *a*. Table 1 shows in:

Column 1: the ratio of aggregate man-hours worked by wage and

\* Eight of the studies covered in my *Unionism and Relative Wages* also examined employment effects of unionism, though in some of them the examination was quite brief: Albert E. Rees, "Postwar Wage Determination in the Basic Steel Industry," *A.E.R.*, June, 1951: steel unionism made both wages and employment in some steel producing centers slightly lower than they otherwise would have been in 1945–47. Irvin Sobel, "Collective Bargaining and Decentralization in the Rubber-Tire Industry," *J.P.E.*, Feb., 1954: unionism in rubber tire manufacturing was one of the factors causing geographic decentralization of the industry and work-sharing in the Akron area. Stephen Sobotka, "Union Influence on Wages: The Construction Industry," *J.P.E.*, Apr., 1953: in 32 large cities in 1939 there was a significant negative correlation between the degree of unionization of skilled building craftsmen and the ratio of the months worked by such workers to the months worked of a base group of relatively weakly unionized workers. Rush V. Greenslade, "The Economic Effects of Collective Bargaining in Bituminous Coal Mining" (unpublished Ph.D. dissertation, University of Chicago, 1952): explores effects of unionism on the number of men employed, average days worked per year, and unemployment in the industry as a whole and among coal-producing states. Elton Rayack, "The Impact of Unionism on Wages in the Men's Clothing Industry, 1911–1956," *Labor Law J.*, Sept., 1958: employment in nonunion producing centers in men's clothing manufacturing gained substantially relative to employment in union centers during the 1920's. Melvin Lurie, "The Measurement of the Effect of Unionization on Wages in the Transit Industry" (unpublished Ph.D. dissertation, University of Chicago, 1958): examines impact of unionism in the transit industry on labor turnover. Milton Friedman and Simon Kuznets, *Income from Independent Professional Practice* (N.B.E.R., 1945): presents evidence on the effects of the American Medical Association on entry of enterprises into the medical training industry and of persons into medical training and practice. Gordon Tullock, *The Sources of Union Gains* (Research Monograph 2, Thomas Jefferson Center for Studies in Political Economy, University of Virginia: Charlottesville, 1959): examines the impact of unionism on the distribution of employment among industries.

salary workers in Group *a* to the corresponding aggregate for Group *b*, with 1929=1

Column 2: the similar ratio for the number of full-time equivalent wage and salary employees, with 1929=1

Column 3: the similar ratio for the average hourly compensation

TABLE 1

EMPLOYMENT, WAGES, NATIONAL INCOME, EXTENT OF UNION MEMBERSHIP, AND  
WAGE EFFECT OF UNIONISM, GROUP *a* RELATIVE TO GROUP *b*

YEAR	EMPLOYMENT (1929=1)		AVERAGE HOURLY COMPEN- SATION <i>W</i> (3)	NATIONAL INCOME (1929=1) <i>Q</i> (4)	EXTENT OF UNION MEMBERSHIP (Group <i>a</i> -Group <i>b</i> )		RELATIVE WAGE EFFECTS OF UNIONISM (COM- MON LOGS)	
	Man- hours <i>E</i> <sub>0</sub> (1)	Em- ployees <i>E</i> <sub>1</sub> (2)			<i>P</i> <sub>1</sub> (5)	<i>P</i> <sub>2</sub> (6)	<i>B</i> <sub>1</sub> (7)	<i>B</i> <sub>2</sub> (8)
1919	1.280	1.218	—	—	—	—	—	—
1920	1.352	1.189	1.416	1.142	0.259	0.286	-0.064	-0.053
1921	1.036	0.995	1.307	0.929	0.269	0.254	0.093	0.082
1922	1.074	1.048	1.158	0.923	0.188	0.161	0.127	0.118
1923	1.162	1.148	1.205	1.046	0.151	0.139	0.102	0.102
1924	1.074	1.066	1.246	0.968	0.161	0.158	0.111	0.108
1925	1.072	1.064	1.200	0.986	0.158	0.156	0.095	0.094
1926	1.049	1.064	1.182	1.008	0.150	0.149	0.091	0.094
1927	1.015	1.014	1.204	0.944	0.157	0.157	0.111	0.109
1928	0.983	0.992	1.199	0.955	0.148	0.144	0.103	0.102
1929	1.000	1.000	1.172	1.000	0.137	0.133	0.101	0.102
1930	0.889	0.938	1.218	0.965	0.145	0.141	0.131	0.122
1931	0.771	0.851	1.231	0.828	0.170	0.169	0.206	0.181
1932	0.692	0.778	1.187	0.701	0.180	0.170	0.258	0.218
1933	0.715	0.809	1.245	0.748	0.170	0.165	0.221	0.182
1934	0.745	0.859	1.299	0.882	0.217	0.243	0.144	0.116
1935	0.777	0.871	1.282	0.869	0.196	0.200	0.138	0.113
1936	0.822	0.895	1.287	0.987	0.202	0.212	0.100	0.082
1937	0.828	0.924	1.362	1.002	0.289	0.333	0.092	0.079
1938	0.709	0.828	1.379	0.849	0.303	0.305	0.102	0.081
1939	0.780	0.872	1.367	0.954	0.304	0.315	0.107	0.087
1940	0.811	0.903	1.389	1.047	0.323	0.344	0.098	0.082
1941	0.968	1.024	1.414	1.253	0.345	0.383	0.049	0.044
1942	1.113	1.132	1.481	1.385	0.364	0.400	-0.009	-0.003
1943	1.231	1.207	1.439	1.442	0.390	0.422	-0.024	-0.013
1944	1.231	1.195	1.395	1.379	0.413	0.428	-0.000	0.010
1945	1.116	1.094	1.344	1.165	0.421	0.413	0.016	0.025
1946	1.019	1.030	1.289	0.983	0.432	0.434	0.005	0.011
1947	1.064	1.057	1.280	1.112	0.456	0.475	-0.028	-0.018
1948	1.053	1.045	1.303	1.146	0.451	0.458	-0.010	-0.002
1949	0.960	0.974	1.317	1.098	0.452	0.439	0.025	0.028
1950	0.994	1.003	1.324	1.198	0.432	0.430	0.045	0.047
1951	1.040	1.034	1.352	1.268	0.456	0.478	0.019	0.026
1952	1.032	1.030	1.361	1.243	0.451	0.458	0.036	0.042
1953	1.048	1.048	1.378	1.294	0.477	0.495	0.048	0.053
1954	0.976	0.985	1.374	1.193	0.460	0.444	0.065	0.065
1955	0.988	0.984	1.374	1.259	0.458	0.459	0.071	0.073
1956	0.981	0.970	1.389	1.254	0.452	0.454	0.064	0.066
1957	0.957	0.945	1.390	1.222	0.465	0.469	0.056	0.059
1958	0.874	0.868	1.392	1.093	0.469	0.454	0.061	0.059

(wages and salaries plus wage supplements) of wage and salary workers<sup>4</sup>

Column 4: the ratio of the national income originating in Group *a* to that originating in Group *b*, with 1929 = 1<sup>5</sup>

Columns 5 and 6: two sets of estimates of the absolute excess (in decimal points) of the degree of unionization of wage and salary employees in Group *a* over that in Group *b*

Columns 7 and 8: two sets of estimates of the impact of unionism on average hourly compensation in Group *a* relative to that in Group *b*, in common logarithms per percentage point difference in the degree of unionization of the two groups

The sources and methods of constructing the series in columns 1-6 are given in Chapter VI of *Unionism and Relative Wages*. The estimates (*B2*) in column 8 were obtained from the regression of the log of the relative wage series (*W* in column 3) on the log of the relative national income series (*Q* in column 4), the rate of unemployment in the labor force, the log of the ratio of the actual to the "expected" price level, degree of unionization (*P2* in column 6), and the mathematical products of *P2* and the unemployment and price-level variables.<sup>6</sup> The series *B1* came from a similar regression with the unionization series *P1* substituted for *P2*. I suspect that, for what they purport to estimate, *B1* and *B2* are too large and fluctuate too widely, but if I could prove my suspicions, I would present different estimates.

Question: To what extent has unionism affected the relative employment series *E0* and *E1*? In answer to this question I have regressed the common logs of *E0* and *E1* on the common logs of *Q*, the common logs of lagged relative employment, and the product of *B* and *P*.<sup>7</sup> The relative demand-supply model underlying these regressions is:

$$\text{demand: } \log E_t = \gamma[a_0 + a_1 \log Q_t - a \log W_t] + (1 - \gamma) \log E_{t-1}$$

$$\text{supply: } \log E_t = \delta[c_0 + c \log W_t - (a + c)B_t P_t] + (1 - \delta) \log E_{t-1}.$$

Eliminate *W<sub>t</sub>* between the equations, obtaining

$$\log E_t = b_0 + b_q \log Q_t + b_s \log E_{t-1} - a(1 - b_s)B_t P_t,$$

where the *b*'s depend upon  $\gamma$ ,  $\delta$ , and the *a*'s and *c*'s.

Table 2 reports the regression results. In both the man-hour (*E0*)

<sup>4</sup> The average hourly compensation figures for Groups *a* and *b* underlying column 3 have fixed weights by industry.

<sup>5</sup> I also computed for the period 1929-58 the index of the ratio of national income plus depreciation originating in Group *a* to that originating in Group *b*. This index is almost the same as that in column 4.

<sup>6</sup> See regression No. 7 in Table 62 of *Unionism and Relative Wages*.

<sup>7</sup> First differences of these variables were used in the regressions.

and employees ( $E1$ ) regressions the regression coefficients for the unionism variable  $BP$  are negative, indicating that unionism tended to reduce employment in Group  $a$  relative to that in Group  $b$ . Moreover, although the unionism coefficients in the man-hour equations were larger numerically than those in the employees equations, none of the unionism coefficients differs significantly from (minus) unity, suggesting that the order of magnitude of the relative employment effects was roughly the same as that of the relative wage effects.

TABLE 2  
MULTIPLE REGRESSIONS FITTED TO THE DATA OF TABLE 1\*

REGRESSION NUMBER	DEPENDENT VARIABLE	REGRESSION COEFFICIENTS AND STANDARD ERRORS				$R^2$
		$\log Q_t$	$\log E_{t-1}$	$B1_t \cdot P1_t$	$B2_t \cdot P2_t$	
1	$\log E0_t$	0.62 (0.07)	0.03 (0.07)	— —	-1.24 (0.33)	0.846
2	$\log E0_t$	0.62 (0.07)	— —	— —	-1.23 (0.33)	0.845
3	$\log E0_t$	0.56 (0.07)	— —	-1.24 (0.31)	— —	0.851
4	$\log E1_t$	0.50 (0.05)	— —	— —	-0.70 (0.23)	0.858
5	$\log E1_t$	0.46 (0.05)	— —	-0.74 (0.22)	— —	0.866

\*  $R^2$  is the square of the multiple correlation coefficient. The standard errors of the regression coefficients are in parentheses.

Let  $b_u$  be the regression coefficient for the unionism variable. Then the antilog of  $b_u B_t P_t$  is an estimate for date  $t$  of the index of the effect of unionism on employment in Group  $a$  relative to employment in Group  $b$ . In 1945-49 the average value of this index differed by 1 percent or less from unity for both man-hours and employees. The values of the index, on a 1945-49 base equal to unity, for more recent periods are (these indexes were computed from regressions 1 and 4 and Table 1):

PERIOD	INDEX OF RELATIVE EMPLOYMENT EFFECT OF UNIONISM (1945-49 = 1)		INDEX OF ACTUAL RELATIVE EMPLOYMENT (1945-49 = 1)	
	Man-hours	Employees	Man-hours	Employees
1945-49	1.000	1.000	1.000	1.000
1950-54	0.951	0.972	0.977	0.981
1955-58	0.929	0.959	0.911	0.906

For man-hours the decline from 1945-49 to 1955-58 in the index of the relative employment effects of unionism is about four-fifths as large as



the corresponding actual decline in relative employment; for the number of full-time equivalent employees, the corresponding decline in the unionism index is approximately four-ninths of the actual decline.

4. The postwar peak in employment (in man-days and man-hours) in bituminous coal mining came in 1947. From 1947 to 1961 the average number of men employed on active mine days declined by almost two-thirds, man-days fell by three-fifths, and man-hours by about three-fourths. Unquestionably much of the decline was the result of falling demand for coal. Yet the presence of the United Mine Workers in the industry<sup>8</sup> and the unusually large postwar climb in wages suggest that unionism may have been a significant factor in the employment decrease.

Table 3 shows index numbers (with 1945=1) of the ratios of average hourly earnings and average hourly compensation for bituminous coal mining wage earners to the corresponding figures for production workers in manufacturing.<sup>9</sup> (Before 1929 the compensation ratios must have been almost the same as the earnings ratios.) In 1945 the ratio of coal to manufacturing hourly earnings was at approximately the same level as in 1890-93 before the advent of collective bargaining in the industry and in 1929-31 when the union was struggling to stay in business in bituminous coal mining. These comparisons suggest that in 1945 the impact of unionism on the relative wage position of bituminous coal miners was close to zero. In 1957 and 1959 the relative average hourly earnings of bituminous coal workers were 28 percent and relative average hourly compensation was 42 percent above the 1945 level. From

<sup>8</sup> I do not have what I consider to be reliable estimates of the extent to which the production and development workers in the industry are covered by collective bargaining agreements. In 1954 slightly more than 20 percent of the workers were employed in counties (or alternatively in establishment-size classes) in which average hourly earnings were less than the minimum rates in UMW contracts. Throughout the period 1947-57 the UMW collected tonnage-based health and welfare fund payments on about 80 percent of the industry output. Since 1957, however, the fraction has declined by about 6 percent, suggesting a decline below 80 percent in the extent of collective bargaining coverage. In this connection notice that, according to Table 3 below, the ratios of bituminous coal to manufacturing average hourly earnings and average hourly compensation have fallen by 8 and 6 percent, respectively, since 1959.

<sup>9</sup> The underlying figures for manufacturing for the years 1890-1957 are from Albert Rees, *Real Wages in Manufacturing, 1890-1914* (Princeton Univ. Press, 1961), p. 33, and his *New Measures of Wage-Earner Compensation in Manufacturing, 1914-57*, Occasional Paper 57 (N.B.E.R., 1960), pp. 3-4. His average hourly earnings series was extended to 1961 by means of the wage and man-hour data in the 1961 *Annual Survey of Manufactures*, p. 13, and from 1961 to 1962 by the Bureau of Labor Statistics average hourly earnings series for manufacturing. Average hourly compensation in manufacturing in 1958-62 was estimated by multiplying average hourly earnings by the ratio of total employee compensation to total wages and salaries in manufacturing, the ratio being computed from Office of Business Economics data. The underlying series for bituminous coal mining for the years 1890-1928 is from Rush V. Greenslade, *op. cit.*, pp. 39-41; for 1929-57 from Ethel B. Jones, "Hours of Work in the United States, 1900-1957" (unpublished Ph.D. dissertation, Univ. of Chicago, 1961), Tables 2 and 3. The Jones series were extended to 1962 by means of the BLS average hourly earnings series for bituminous coal mining and the OBE series for employee compensation and wages and salaries.

1959 to 1962 the relative wage position of bituminous coal miners declined by about 10 percentage points.

TABLE 3

INDEXES OF AVERAGE HOURLY EARNINGS AND AVERAGE HOURLY COMPENSATION,  
BITUMINOUS COAL MINING RELATIVE TO ALL MANUFACTURING INDUSTRIES  
(1945=1)

YEAR	AVERAGE HOURLY EARNINGS	YEAR	AVERAGE HOURLY	
			Earnings	Compensation
1890	1.01	1929	0.99	1.01
1891	0.95	1930	0.98	1.01
1892	1.00	1931	0.98	1.00
1893	1.01	1932	0.91	0.94
1894	1.00	1933	0.86	0.88
1895	0.93	1934	0.97	0.99
1896	0.83	1935	1.05	1.07
1897	0.80	1936	1.09	1.11
1898	1.01	1937	1.06	1.08
1899	1.03	1938	1.11	1.12
1900	1.15	1939	1.07	1.08
1901	1.22	1940	1.02	1.03
1902	1.20	1941	1.04	1.06
1903	1.27	1942	0.95	0.97
1904	1.29	1943	0.92	0.93
1905	1.28	1944	0.94	0.95
1906	1.26	1945	1.00	1.00
1907	1.19	1946	1.07	1.08
1908	1.25	1947	1.12	1.15
1909	1.22	1948	1.23	1.29
1910	1.18	1949	1.22	1.29
1911	1.18	1950	1.21	1.32
1912	1.21	1951	1.22	1.30
1913	1.13	1952	1.19	1.28
1914	1.17	1953	1.22	1.34
1915	1.19	1954	1.21	1.34
1916	1.16	1955	1.20	1.36
1917	1.24	1956	1.26	1.40
1918	1.16	1957	1.28	1.42
1919	1.19	1958	1.23	1.37
1920	1.38	1959	1.28	1.42
1921	1.53	1960	1.26	1.40
1922	1.63	1961	1.22	1.36
1923	1.50	1962	1.18	1.33
1924	1.27			
1925	1.15			
1926	1.14			
1927	1.08			
1928	1.05			

These figures may understate the impact of unionism on the average relative wage in bituminous coal mining since they do not take into account the relative wage impact of unionism in manufacturing. Adjustment for the latter would raise the figures in Table 3 in recent years by roughly 4 to 7 percentage points. Hence I estimate tentatively that in bituminous coal mining in 1957-59 the effect of unionism on average

hourly earnings was about 30 to 35 percent and on average hourly compensation about 45 to 50 percent; the corresponding estimates for 1962 are about 10 percentage points lower.<sup>10</sup>

For purposes of estimation, I have divided the employment effect of unionism in bituminous coal mining into three components:

$$\begin{aligned} \text{welfare contribution component: } & \frac{T}{p - T} \eta \\ \text{scale component: } & Rk\eta \\ \text{substitution component: } & R(1 - k)\sigma \end{aligned}$$

where  $T$  is the average contribution per ton,  $p$  is the price per ton at the mines,  $\eta$  is the elasticity of demand for coal at the mines,  $R$  is the relative wage effect,  $k$  is "labor's share" (the ratio of employee compensation less welfare fund contributions to value of output), and  $\sigma$  is the elasticity of substitution of capital for labor. In recent years the value of  $T/(p-T)$  has been about 0.1 and  $k$  approximately 0.5.<sup>11</sup> With  $R$  equal to about one-third in 1957-59, the corresponding values of  $Rk$  and  $R(1-k)$  are about one-sixth.

I am in the process of estimating the elasticity of demand  $\eta$  and do not yet have any numbers to report. But unless  $\eta$  is quite small, the welfare contribution and scale components taken together will not be negligible. For example, even if  $\eta$  were only 0.25, these two components would total about 6 percent.

Although I am not yet satisfied that I have really estimated the elasticity of substitution  $\sigma$ , I do have some results to report. In making the estimates, I have used the now familiar CES (constant elasticity of substitution) productivity-wage relation

$$(1) \quad z = \sigma \log a + (1 - \sigma) \log b + \sigma y$$

where  $z$  is the log of bituminous coal output per unit of coal labor input,  $y$  is the log of the ratio of the coal mining average wage to the price of coal,  $a$  is the "distribution" parameter in the CES production function, and  $b$  the "efficiency" parameter.

I first fitted equation (1) to cross-section (by states) data on  $y$  and  $z$  for 1954 and 1958, separately for strip and underground mines, com-

<sup>10</sup> To check these figures, I am now preparing alternative estimates for each of the years 1947-61 from a cross-section (by states) analysis of wage levels and wage changes in bituminous coal mining.

<sup>11</sup> In 1958 welfare contributions amounted to 10 percent of value added (less welfare contributions) and 7 percent of the value of net coal shipments (less contributions). In the same year, employee compensation (less welfare contributions) of production and development workers was about 50 percent of value added and about 40 percent of the value of net coal shipments.

puted from the 1954 and 1958 *Census of Mineral Industries*, obtaining the following estimates of  $\sigma$  (from the best-fitting regressions):

	REGRESSIONS OF $z$ ON $y$		REGRESSIONS OF $y$ ON $z$	
	Strip	Underground	Strip	Underground
1954	1.05	1.19	1.41	1.29
1958	1.14	1.16	1.35	1.28

These results suggested that  $\sigma$  was approximately the same in strip and underground mining. Therefore, in later work I combined the two types of mines. This made it possible to use the annual data for each of the twenty-two leading coal producing states (including Alaska) on output, price, wages, and labor input provided by the Bureau of Mines (*Minerals Yearbook*) and the Bureau of Employment Security. However, since the level of man-hour productivity, state by state, is substantially higher in strip than in underground mines, I have usually included the ratio,  $\lambda$ , of man-days worked in underground mines to total man-days worked as an independent variable in the regressions.

For each year 1947 to 1961 I have computed the regressions across the 22 states of  $z$  on  $y$  and  $\lambda$  and of  $y$  on  $z$  and  $\lambda$ . The estimates of  $\sigma$  obtained from the regressions, though varying somewhat from year to year, have no trend to speak of over the period. The 1947-61 averages of the estimates of  $\sigma$  are:

regressions of  $z$  on  $y$  and  $\lambda$ : 0.96  
regressions of  $y$  on  $z$  and  $\lambda$ : 1.31.

Since geological conditions (and perhaps labor quality as well) probably varied more among the states at given dates than among the years 1947-61 within states, estimates of  $\sigma$  derived from time series may be preferred to those from the annual geographic cross sections. Hence for each of the twenty-two states I also computed estimates of  $\sigma$  from the following 8 regressions:

1.  $z_t$  on  $y_t$ ,  $z_{t-1}$ ,  $\lambda_t$ , and time,  $t$ . . . . . 0.95
2.  $z_t$  on  $y_t$ ,  $z_{t-1}$ , and time,  $t$ . . . . . 0.92
3.  $z_t$  on  $y_t$ ,  $z_{t-1}$ , and  $\lambda_t$ . . . . . 1.25
4.  $z_t$  on  $y_t$  and  $z_{t-1}$ . . . . . 1.40
5.  $y_t$  on  $z_t$ ,  $z_{t-1}$ ,  $\lambda_t$ , and  $t$ . . . . . 1.40
6.  $y_t$  on  $z_t$ ,  $z_{t-1}$ , and  $t$ . . . . . 1.50
7.  $y_t$  on  $z_t$ ,  $z_{t-1}$ , and  $\lambda_t$ . . . . . 1.34
8.  $y_t$  on  $z_t$  and  $z_{t-1}$ . . . . . 1.41.

There was considerable variation among the estimates from state to state and in eight states (producing a bit less than one-fourth of the U.S.

bituminous coal output in 1955) the estimates varied so much among the regressions as to be essentially useless. The numbers at the extreme right of the regressions above are weighted averages (the estimate for each state was weighted by man-days worked by coal miners in the state in 1955) of the estimates of  $\sigma$  for the other fourteen states. These time series figures agree rather well with the cross-section estimates.

The substitution component of the employment effect is  $R(1-k)\sigma$ . With  $R(1-k)$  at about one-sixth and  $\sigma$  at 0.9 to 1.5, the substitution component alone amounts to 15 to 25 percent, an effect about two to three times as large as the total employment effect estimated for the late fifties in the global data for industry Groups *a* and *b*. On the other hand, unless the elasticity of demand for coal at the mine turns out to be substantially larger than unity, the total effect (substitution plus scale and welfare contribution components) of unionism on employment in bituminous coal mining will account for considerably less than one-half (in ratio terms) of the postwar employment decline in that industry.

# LABOR MOBILITY AND WAGE INFLEXIBILITY

By ROBERT L. RAIMON  
*Cornell University*

## I

However useful would be an increase in wage flexibility to the prospects for low unemployment at stable prices, we seem to lack a practical way of achieving it. So long as this be true, efforts to improve the quality of labor mobility, through training and relocation programs, may be expected to attract increasing support. Indeed, should it become apparent that increases in aggregate demand have to be pushed to inflationary levels to reduce unemployment to an acceptable rate, raising the personal productivity of overpriced workers is the only approach to full employment at stable prices that remains open. This, in any event, is the positive side of these remarks.

On the negative side, given the considerable evidence of the absence of wage flexibility, is the dependence of any market on price flexibility for efficiency of performance. Market efficiency is measured by the swiftness and appropriateness with which price adjustments establish market-clearing conditions for demand and supply. Efficiency in the labor market involves the relationship of the average level of wages to other factor prices, the flexibility and correctness of response of wage adjustments for particular categories of labor to changing market conditions, and quantity adjustments on the supply side among related sections of the market. Hence it follows that anticompetitive institutional pressures on wages, such as may be introduced by collective bargaining, impair one of the labor market's devices for inducing efficient allocation of workers. Accordingly, the burden increases on quantity adjustments—labor mobility—in each of its dimensions. Further, the wrong kinds of changes in relative wages—those which are anticompetitive instead of market-clearing—impose a double constraint on the quantity-adjustment mechanism: (1) they erect a wage hurdle with a limiting effect upon sectional expansion in employment or they make sectional employment contractions larger than they otherwise would be and (2) they deepen worker attachments where demand conditions are unfavorable.

Perspective on some of the problems in the present-day labor market is gained by noting first that the shift away from employment of production workers in the goods sector imposes a strain on wage differ-

ences as an allocative device. In this connection, reference to the contrast between American and foreign experience is helpful.

## II

Comparing the high labor mobility and minimal job tenure traditions found in the United States with the low turnover rates that typically prevail in Europe and Japan, it is easy to doubt the decisiveness of fluidity in the labor market to rapid growth and full employment. While foreign experience suggests that if other markets are in good repair an economy can successfully accommodate a wide variety of obstacles to the efficient matching of workers and jobs, the Europeans and the Japanese seem to be sailing with the wind behind them. We, at a different stage of development, have to zig and zag. During the past decade, in virtually all industrialized countries except ours, manufacturing employment has been growing more rapidly than the labor force [23]. Since manufacturing generally is a high-wage sector, the path to improved allocation of labor supplies is smoothed as attractive job opportunities multiply. To be sure, all economies confront structural problems. As far as the functioning of our labor market is concerned, performance is not likely to be enhanced by the restraints which have come to prevail in the blue-collar ranks of American industry. There, the laid-off worker without the credentials to qualify for a white-collar job often finds himself confronted with the choice between awaiting recall or seeking employment in one of the expanding service occupations.<sup>1</sup> Since the latter are generally lower paid, less skilled, and often part time, it is not surprising that in recent years between one-third and one-half of total accessions in manufacturing have been recalls. If the observations could be limited to production workers, the proportion would be considerably higher. The higher the industry wage level, the greater the tendency for recalls to outnumber new hires.

Our industrial labor markets are somewhat akin to the farm problem. Even with the help of a favorable earnings spread, the farm population shrinks too slowly. Now we seem to require a relative contraction in the hourly-rated portion of the goods-producing industries, one that inevitably will entail many direct, individual worker transfers, but the relative wage signal is pointing in the wrong direction.

<sup>1</sup>The service occupations should not be confused with the service industries. In the former group, with the exception of a few occupations, like policeman or bartender, virtually all of the jobs are unskilled. The fuller employment is expected to be, the less relative shrinkage will occur in the demand for goods-producing production workers. The BLS, on the basis of a 3 percent unemployment rate assumption, projects a 1960-70 increase of 21 percent in all employment, including a gain of 13 percent (1,600,000) in operatives and a 34 percent (2,800,000) in service workers [26, Table 28, p. 100].

## III

Despite its much described imperfections, the American labor market scores high, both when measured on an impressionistic basis against European experience and when subjected to formal tests, in the sense of satisfying certain of the predictive implications of the competitive hypothesis, particularly those pertaining to geographic mobility. See, for example [5]. Nevertheless, its performance is disappointing—especially with respect to the level of employment; that is, as has been suggested, certain aspects of the labor market itself contribute to the economy's almost chronic failure to generate full employment.

Predicting the tolerance of a society for unemployment is impossible. Leo Wolman used to say that unemployment was a problem when people worried about it, and perhaps that is as precise a statement as one can make. We know that our tolerance level is a moving target and that it is moving downwards. Among the forces that may continue the pressure to reduce unemployment is the pace set by other industrialized countries. Also, we are likely to experience persistent and perhaps increasing concern with the labor market aspects of the civil rights movement and with the concentration of unemployment among the rising numbers of young persons in the labor force. Especially dim are the job prospects for the several million high school dropouts who are entering the labor force this decade.

Basic to the current surge of interest in labor mobility and manpower policies in connection with the high unemployment of the past six years is the suspicion that somehow a serious imbalance may have emerged between the skills the unemployed have to offer and the kinds of skills that employers seek to hire at prevailing and prospective wage rates. Mention of the notion of a bad skill match between labor demands and supplies plunges one into the inadequate demand-structural distortion controversy. Without respect to the relative merits of this dispute at unemployment rates well above 5 percent, there is concern that as we approach 4 percent further expansion of aggregate demand is likely to exert more pressure on prices than on employment, in part because of constraints imposed by the labor market. Among these, three stand out:

1. The record of the 1950's suggests that some unions are strong enough to command substantial wage increases while the companies with which they bargain still face perfectly elastic supply curves of labor at prevailing wage levels. According to a familiar line of reasoning, such a development, assuming continued priority for the price stability objective, prompts restrictionist monetary and fiscal policies if the cost increases exceed the limits of productivity progress. A wage



spillover into the nonunion sector is a part of the process. The end result is a continuing unemployment problem. Further, as Abba Lerner has suggested, government efforts against sellers' inflation may be awkward: despite a tight money policy, prices may keep rising until real expenditures and employment have been reduced sufficiently to overcome the institutional forces that confer upon sellers the power to keep raising prices [10].

One finds in the labor market studies of the late 1940's and early 1950's strong support for this anticompetitive institutional control of the labor market, coupled naturally to a denial of the relevance of conventional labor market theory to labor market analysis. This interpretation was best summarized by Arthur M. Ross: "Relative wages have little to do with the movement of labor, and . . . the movement of labor has little to do with the creation of wage differences" [16]. If cost-push tendencies keep full employment at too respectful a distance, then efforts to improve the quality of labor mobility are unlikely to prove fruitful. The development of labor shortages in some sectors simultaneous with surpluses in others is improbable on a large scale with unemployment high almost everywhere. To be sure, shortages may arise for workers in the upper echelons of the occupational hierarchy, but these will tend to be self-correcting. Individuals may be expected to perceive that the chief escape from the inhospitable parts of the labor market lies in qualifying for those skills for which employers are freer to respond to low bids.

2. It may be objected that even in the absence of cost-push, a level of demand sufficient to cause full employment for skilled workers may be inadequate to reduce unemployment below a tolerable level for lesser skilled workers [7] [8]. That the incidence of unemployment is suffered disproportionately by those at the low end of the skill and education ladder is firmly established. Why this should be so is not clear. The growth of research concerning investment in human capital promises to improve our understanding of this phenomenon.

3. Pending the appearance of a better theory directed to the apparently increasing concentration of nonrecession unemployment among those with low personal productivity, Clarence D. Long's work commands the center of the stage [12] [13]. Briefly put, Long points to the operation of two forces: first, a "social minimum wage" that is rising as fast or faster than the average wage; second, as improving educational facilities enhance "the opportunity offered to the average worker to improve his personal productivity, the further some of the

\*The social minimum wage is a composite of legal and union minima, employer reluctance to pay low wages, as well as the alternatives to wage income found in the social and private insurance systems.

workers will fall below the average. That is, the faster the march of the average, the greater the lag for the stragglers" [12]. This is the widening productivity spread. Thus the productivity of the average worker is rising more rapidly than the productivity of those who for whatever combination of environmental and genetic reasons fail to move as fast as the average. Meanwhile, because of the upward drift of the social minimum wage, no compensatory widening of the wage spread between the two groups is developing to offset the widening productivity spread.

A major virtue of the Long hypothesis is that it introduces, through relative wages, an economic explanation for the failure of the skill needs of employers to match those that the unemployed have to offer. That is, employers have responded to the relative rise in the cost of low quality labor by raising hiring specifications. That the available jobs often have a new look, that processes have been automated, simply obscures the underlying sequence of cause and effect. Although there may be occasions when it pays to automate even at zero wage rates, change in applied technology is usually influenced by change in factor costs, including the cost of labor. Also, because the long-run supply of capital is more elastic than that of labor, the price of labor rises relative to that for capital. This interfactor competitive process induces a substitution against labor which may prompt the disemployment of overpriced workers. In short, inappropriate wage behavior puts an uneconomic premium on laborsaving innovations.

#### IV

Common to the Long theory of prosperity unemployment and to the threat of cost-push to sustained high full employment, is our wage-making system's deficient sensitivity to the imbalance between demand for and supply of labor in important sectors. Certainly there is nothing novel about the notion that wage flexibility conduces to full employment, provided it does not trigger a deflationary spiral. Nor is such wage behavior a purely hypothetical construct. Wage decisions for manual workers are less responsive to labor market conditions than are those for white-collar workers. This conclusion, at least, is indicated by a comparison of the New Haven study of manual workers [15] with the study of female clerical workers in Boston banks and insurance companies [18].

The growing literature about the Phillips curve seems to suggest that wages have tended to rise faster at given levels of unemployment as collective bargaining has spread. Between 1947 and 1961, despite the difficulties of moving from blue- to white-collar employment, disbursements per salaried worker in manufacturing increased at an average

annual rate of 3.8 percent while the comparable rate for production workers was 4.9 percent. During these years the number of salaried workers was rising at a rate of 3.7 percent a year and the number of production workers was falling at a rate of 0.5 percent a year [24]. Similarly, between 1953 and 1962, among twenty large cities for which information about straight-time average hourly earnings is available, median wages for both unskilled male plant workers and skilled maintenance workers alike increased by 47 percent [22]. During these years, unemployment rates for experienced male unskilled workers were two to three times the rates for skilled workers, averaging about 11 percent. However, this level was exceeded by the unemployment experience of male youths, age 18-19, many of whom presumably were candidates for unskilled work. Also, except for the textile centers, one finds no tendency for manufacturing wage levels to rise least where labor supplies are locally most abundant [17]. Average hourly earnings in depressed West Virginia have more than kept pace with wage changes in manufacturing as a whole.<sup>3</sup> The general failure in the articulation of manufacturing wage changes to local conditions appears to have the kind of related consequences one would expect. One study points to an apparent wage spillover from manufacturing into the service industries [2]. The price impact of such an effect probably reduces the employment absorption capacity of the service sector.

In summary, the responsiveness of our wage-making system for blue-collar workers to the conditions of supply and demand for such workers, as so often has been noted, frustrates high full employment. In contrast to this nonadaptive wage behavior, what appears to be adaptive movement is much in evidence when one turns to information descriptive of worker mobility. That is, wages rarely move downwards but workers often do.

## V

Some sense of the flexibility of the work force is provided by the observation that the most characteristic kind of job change (exclusive of changes that take place within the firm, about which we know little) is one that involves the simultaneous change of employer, occupation, and industry. This pattern is reported for 1940-50 by the Six City project [14] and repeated by two nationwide studies of job mobility, one pertaining to 1955 [19] and one to 1961 [1]. More than one-half of the job shifts reported by the Six City study were of this multiple character,

<sup>3</sup>The reference is to a comparison of 1951, the first year for which data are available, with 1962. West Virginia wage data are in [21]. The results are not the consequence of any important change in industry-mix; however, changes in the skill-mix were not investigated.

almost one-half in 1955, and well over one-third in 1961. (No inference is justified as to a rise in occupational or industrial attachment, however, in part because of differences in labor market conditions associated with the timing of each of the surveys.) Recognition of the high degree of flexibility in worker movement that these findings convey is enhanced by noting that the occupational and industrial categories used are of the gross kind found in the decennial census for classifying occupations and industries into a dozen or so groups. Self-evidently, a more finely graded classification system would further reduce the occupational and industrial attachment that the labor force demonstrates.

With respect to geographic mobility, the sheer volume and frequency of movement is impressive. Between 1955 and 1960, almost one-half the population, age five years and over, changed residence. During the 1950's, four out of five counties lost residents as a result of net out-migration [26, pp. 55, 56]. New research indicates that half of the 800,000 workers who change jobs each month find their new work in a geographically different labor market [27, p. 31].

Judging from the recovery period extending from the summer of 1954 through the summer of 1957, there is sufficient mobility, in all of its dimensions, including into and out of the labor force, to overcome all but occasional shortages of general labor supplies in particular localities while surpluses obtain elsewhere. During this period, the Bureau of Employment Security estimated labor market conditions for each of 150 major "production and employment centers" on a bimonthly basis [20]. In most instances boundaries of these areas coincide with those of the Standard Metropolitan Areas. Each area was categorized by one of six designations descriptive of the tightness or looseness of its labor supplies. Areas assigned to the top category are estimated to be experiencing "overall labor shortages" associated with a "usual unemployment rate" of less than 1.5 percent.<sup>4</sup>

Of the 2,700 observations recorded during the recovery, only 10 were of the shortage-designating variety. The national unemployment rate averaged 4.4 percent during this period of concern over wage-push. How efficient the labor market will prove to be when confronted with a more severe test remains to be seen—both with respect to the evenness of the distribution of unemployment and to the pressure exerted on the wage level while unemployment remains a problem.

Turning to skilled and professional workers, the Commissioner of Labor Statistics complains that shortages of such talent occur in one geographical area while an oversupply of the same kinds of workers prevails in other areas [3]. Presumably such observations are based on

<sup>4</sup> A somewhat different scale was in use in the early part of the recovery, but the difference is of little importance to a cross-sectional view.

the interarea recruitment program of the public employment service. In the absence of a job vacancy series and more complete use of the employment service by high quality workers, the direct evidence available is insufficient to support a judgment on the magnitude of this problem. Indirect evidence, such as might be found in the relative movement of wages by occupation and place, is becoming more abundant thanks to the Bureau of Labor Statistics' expanded occupational wage survey program.

The mobility studies agree that industry attachment tends to be disguised occupational attachment. It is change in the occupational more than in the industrial structure of employment that most strains the adaptive capacities of the labor force. In general, however high may be the propensity of workers to change occupations and industries, the concept of noncompeting groups appears to have lost none of its relevance. There are no estimates of the number of job vacancies nor of the number that would remain unfilled if worker ignorance and transfer costs were assumed away. Seymour Harris guesses that "adequate manpower training, area redevelopment and vocational programs might cut unfilled vacancies by at least 500,000, thus reducing excess unemployment by at least one-third" [6]. The present remedial programs fall very short of that order of magnitude. Moreover, because of the suspected large number of low-skilled workers and women who only await a more attractive labor market to enter the labor force, estimating the need for training programs is especially difficult.

## VI

To recapitulate, if the labor market is better to serve the objectives of full employment at stable prices, including full employment for nonwhites and the millions of ill-prepared youths who are entering the labor force, progress can come either from increasing wage flexibility, from improving worker mobility—especially by increasing the employability of unskilled workers—or from both together.

Given the downward inflexibility of wages and prices, it is the task of monetary and fiscal policy to lower the unemployment rate from the 5 to 6 percent level prevailing since 1958. But as the Wage Guideposts statement attests, our wage-making arrangements may threaten price stability before we attain an acceptable level of employment, and thereby make more difficult further reduction in unemployment. Similarly, despite the concentration of unemployment among low-skilled workers, there is but limited evidence of relief appearing in the form of a widening of the wage rate spread between less and more skilled work. Such evidence as there is takes two forms: First, the long-run trend toward compression seems to have been arrested since the early 1950's.

Second, there has been an interindustry shift—away from the employment of operatives and laborers in the goods sector and toward the low-wage service industries. Unfortunately for unskilled males, however, the bulk of the rise in such employment has been for women. As Long suggests, the women often are better educated and accept lower wages.

Annual earnings of skilled workers have increased more rapidly than earnings of unskilled workers since 1950, but this seems to be a reflection chiefly of differential exposure to unemployment. Also, it is to be noted that the 1961 amendments to the Wage and Hour Law will not cease pushing up the wage floor until September, 1965.

Similarly, it is not realistic to expect an increase in wage flexibility in collective bargaining of a type that would stress wage structure rather than general wage changes. True, high unemployment does dampen the rate of general increases. But while this is one type of wage flexibility, it is not structural flexibility. Unfortunately, it is all that we can probably expect but not enough to make a positive contribution toward achieving full employment. Wage adjustments tailored to the labor supply and demand situation of individual establishments are the kind that would be most effective. The postwar record indicates that as far as the wages of production workers are concerned, often we do not even find such tailoring at the industry level. There have been repeated observations of an absence of short-run positive correlation between wage and employment changes by industry. Instead, profits [4] and wage parity are found to be among the major elements that condition union wage demands and nonunion wage expectations. Neither, needless to say, necessarily lends to the labor market those qualities that make for low unemployment or equalization of net attractiveness among jobs.

## VII

The proposals that have been advanced to improve efficiency in the labor market are varied and often wonderful to behold. At one extreme is the judgment that if improvement is to come, it can only come from the unions committing *hara-kiri*—capitalism and unionism cannot co-exist. At the other end of the scale are the champions of unrestrained “free collective bargaining,” although even in this camp one suspects an awareness of the inevitability of continued assertion of the public interest in wage decisions.

Generally speaking, aside from occasional urgings to strengthen the placement service, most of the proposals, not surprisingly, have dealt with the pricing of labor, either directly or through suggestions to modify public policy toward collective bargaining. Of late, there has been a marked increase in emphasis on training and relocation.

In terms of the actual adoption of wage policy, as everyone knows, from occasional appeals for restraint we have moved to admonition with guideposts; in effect, to a policy of appealing to the parties who enjoy market power more or less to simulate the behavior they would be obliged to pursue if they did not have market power. Thus far, the guideposts have been ignored in the transportation industry and most notably in construction, despite the prevalence of substantial unemployment in the building trades. The prescription that average wage gains not exceed the rate of secular overall improvement in output per man-hour means that some wages should increase less than others. In those industries where workers are in excess supply, below average increases are called for. But for this to occur, there has to be an appropriately distributed supply of patsies. In the face of the current and projected shift in the distribution of employment away from the unionized sectors, the relatively high wages characteristic of unionized industries, and the unlikelihood in the next several years of a dearth of job applicants at these wages, the guidepost policy seems destined not to enjoy a fortuitous congruence of market power and market forces.

If the guideposts fail while the need continues for a substitute less onerous than high unemployment, attention may return to some of the other proposals that have been put forth. These include advance notice of impending wage and price changes; subjecting "key" industries to utility-type controls; additions to the list of proscribed bad practices with respect to the exercise of union power; direct wage and price controls; subjecting unions to antitrust—whatever that may mean; and reducing the scale of collective bargaining to the enterprise level. For the latter proposal to work, as its author, H. Gregg Lewis indicated, large firms must be made into smaller firms to whatever degree is necessary in order to cope with the problem of follow-the-leader collusion in concentrated industries [11].

Among still additional proposals, one would refer the wage determination process to a system of universal job evaluation [15, pp. 263-66]. Another, a refined version of the guideposts, contemplates a system of labor market indicators that would reveal the balance of supply and demand in each occupational market; by applying an appropriate wage change formula to these indicators, optimal allocation will result [9].

The foregoing list of proposals does not exhaust those that have been advanced to improve the workings of the labor market through modifying our wage-making institutions. Nevertheless, it is perhaps sufficient to indicate that the kinds of proposals that are consistent with decentralization and competitive pricing are remote of realization. Foreign experience confirms this impression.

However, in the past few years interest in wage determination has

been rivaled by a concern with the structure of unemployment and investment in man. The fruits of these developments are illustrated by the Manpower Development and Training Act and by such research as is reported in the *Investment in Human Beings* volume [25]. The question thus arises to what extent economic analysis, newly expanded to embrace investment in training, in migration, and so forth, may be able to throw light on the possibilities of offsetting, by means of such investments, the mischief that insufficient wage flexibility imparts to the labor market and to the economy. For example, just as workers who moonlight or work long hours at one job do not necessarily reduce job opportunities for the unemployed, so similarly is it likely that not all unemployment is equally effective in holding wage changes to a noninflationary pace. Nudging the Phillips curve to the left may be expected to require a mixture of monetary policy and remedial manpower programs. The underdevelopment of our mobility-improving policies may first and foremost result from the greater ease of ignoring chronic looseness in the labor market than chronic tightness. It is of interest to note in this connection that Sweden is both less inhibited than we are about the application of Keynesian prescriptions and also is geared to accommodate as much as 1 per cent of her labor force in retraining programs [27, p. 966].

Prodded by a concern with the hostile nature of the labor market for Negroes and noncollege youth, public support for training programs may be expected to widen and deepen. Also, while the United States may have exceeded most other industrialized nations in arming her labor movement with legal aids to win wage increases at high levels of unemployment, we seem least prepared to accept the possible price-level consequences of that handiwork—probably even without respect to the balance-of-payments problem. At the same time, we have done comparatively little to relieve the alternative result, unemployment, as is apparent when our pre-employment training, apprenticeship, and relocation efforts are measured against certain foreign programs. Research that will help estimate the potential of improving mobility as an offset to insufficient wage flexibility is likely to be one of the things economists will be doing more of in the years ahead.

That the offset may be only partial inheres in the logic of the market, in the sense that efforts to improve quantity adjustments are always in danger of being neutralized in part or in whole by perverse price adjustments. But while price and quantity adjustments may claim equal importance in allocation theory, they are not equally susceptible to efforts to raise the efficiency of the labor market. The more the concern of the government with wage decisions, the closer we come to the inefficiencies of wage and price controls. However, training programs which



increase the productivity and hence the employment prospects for individual workers need involve no such liabilities.

## REFERENCES

1. G. Bancroft and S. Garfinkle, "Job Mobility in 1961," *Special Labor Force Report* No. 35, Bureau of Labor Statistics, 1963.
2. G. W. Bleile, "The Service Sector: Data on Output, Employment, Prices, and Income," in *Study of Employment, Growth, and Price Levels*, Study Paper No. 21, Joint Economic Committee Print, Washington, 1960, pp. 115-22.
3. E. Clague, "The Future Shape of the Labor Force," an address to the National Industrial Conference Board, New York City, May 16, 1963, mimeo.
4. O. Eckstein and T. A. Wilson, "The Determination of Money Wages in American Industry," *Q.J.E.*, Aug., 1962, pp. 379-414.
5. L. E. Gallaway, "Labor Mobility, Resource Allocation, and Structural Unemployment," *A.E.R.*, Sept., 1963, pp. 694-716.
6. S. Harris, *The New York Times Book Review*, Oct. 20, 1963, p. 3.
7. J. R. Hicks, "Economic Foundations of Wage Policy," reprinted in W. Galenson and S. M. Lipset, *Labor and Trade Unionism* (New York, 1960), p. 186.
8. C. C. Killingsworth, "Automation, Jobs and Manpower," Statement before Subcommittee on Employment and Manpower, Committee on Labor and Public Welfare, 88th Cong., 1st sess., Sept. 20, 1963.
9. A. P. Lerner, *Economics of Employment* (New York, 1951), pp. 209-41.
10. ———, "Inflationary Depression and the Regulation of Administered Prices," U.S. Cong., Joint Economic Committee, *Relationship of Prices to Economic Stability and Growth*, Compendium, 85th Cong., 2nd sess., Washington, 1958, p. 264.
11. H. G. Lewis, "The Labor Monopoly Problem: A Positive Program," *J.P.E.*, Aug., 1951, p. 280.
12. C. D. Long, "A Theory of Creeping Unemployment and Labor Force Replacement," Catholic Economic Assn., Annual Meeting, Dec. 27, 1960, reprinted in 107 *Cong. Rec.* 12455 (daily ed. July 25, 1961).
13. ———, "An Overview of Postwar Labor Market Developments," *Proceedings of the Fourth Annual Social Security Conference*, Kalamazoo, W. E. Upjohn Institute, 1962, pp. 9-23.
14. G. L. Palmer and C. P. Brainerd, *Labor Mobility in Six Cities* (New York, 1954), pp. 62-80.
15. L. G. Reynolds, *Structure of Labor Markets* (New York, 1951).
16. A. M. Ross, Indus. Rel. Research Assn., *Proceedings of the Sixth Annual Meeting*, 1953, p. 267.
17. P. Ross, "Labor Market Behavior and the Relationship Between Unemployment and Wages," Indus. Rel. Research Assn., *Proceedings of the Fourteenth Annual Meeting*, 1961, pp. 275-88.
18. G. P. Shultz, "A Nonunion Market for White Collar Labor," in National Bureau of Economic Research, *Aspects of Labor Economics* (New York, 1962), pp. 107-46.
19. Bureau of the Census, *Current Population Reports*, Labor Force, Series P-50, No. 70, "Job Mobility of Workers in 1955," Feb., 1957.
20. Bureau of Employment Security, *Bimonthly Summary of Labor Market Developments in Major Areas and Area Labor Market Trends*.
21. Bureau of Labor Statistics, *Employment and Earnings Statistics for States and Areas, 1939-62*, Bull. No. 1370, 1963.
22. ———, *Wages and Related Benefits, 1961-62*, Bull. No. 1303-83, p. 49 and Table 4, p. 54.
23. *Comparative Wage Costs and Trade Advantage*, House Committee on Education and Labor, Committee Print, 88th Cong., 1st sess., Jan., 1963, Tables 46-47, pp. 86-87.
24. *Economic Report of the President*, Jan., 1962 (Washington, 1962), p. 176.
25. *J.P.E.*, Supplement, Oct., 1962.
26. *Manpower Report of the President* (Washington, 1963).
27. *Nation's Manpower Revolution*, Hearings before the Subcommittee on Employment and Manpower, Committee on Labor and Public Welfare, 88th Cong., 1st sess., May-June, 1963.

## THE FEATHERBEDDING PROBLEM\*

By PAUL A. WEINSTEIN  
*Columbia University*

A growing number of disputes appear to involve the problem of featherbedding. At the heart of these conflicts is the question of who should bear the cost of technological change. While there are a number of theoretical alternatives to technological displacement, featherbedding is the most satisfactory from the position of both the union and the potentially displaced workers. There has been increasing pressure on unions not to resort to featherbedding demands, but external pressure, even when directed by the President, has not proved effective. Some collective bargaining agreements have offered alternatives to featherbedding when there is a structural change, but the advice to follow the lead of constructive agreements that propose to end constraints on firms and effect a better utilization of our manpower resources has gone unheeded. The failure to settle featherbedding issues has cast doubt upon collective bargaining and the market as an appropriate institution.

This paper will treat the following topics: first, the necessary and sufficient conditions for the emergence and continuance of featherbedding; second, the impact of the featherbedding rules on decision making in both the short and long run; third, alternatives to featherbedding for meeting structural unemployment.

### *I. Conditions for Featherbedding*

Employment insecurity gives rise to featherbedding under special circumstances. The following generalizations are significant in understanding the acceptance or rejection of this particular response to the problem of insecurity.

Featherbedding occurs in industries that are characterized by non-competitive operation prior to formalization of the rules. The economic environment for featherbedding is very similar to that required for racketeering. In both cases the firms must have some expropriatable surplus, or else the working rule or extortion leads to downward instability. In the main, the distinction between these two phenomena is in the utility functions of the maximizing institution and the division of the rewards. Therefore, it is inappropriate to examine the im-

\*I wish to thank D. Brody, D. Dewey, and A. Warner for their comments. This paper was prepared with the help of a grant from the Research Council in the Social Sciences of Columbia University.

pact<sup>1</sup> of the rules on the assumption that they disturb an optimal competitive position.

For example: the industries protesting most about these rules—and where in fact they are found in abundance—are the transportation industries and the building trades. Both are noncompetitive in character. The railroads, for example, are under extreme control in nearly every activity, from quality of product to price, while the building trades continue to be a hotbed of localized monopoly in which the government has a role through licensure and demand. Given this state of affairs, some might suggest we dismiss the problem merely as one of internal allocation of quasi-rents. Featherbedding ties up manpower in unproductive activities, and to dismiss the problem as one internal to the firm leaves the larger manpower problem unresolved. It is better to have a higher utilization of human resources even if there is monopoly, though quite obviously it is best to free both labor and product markets.

The spread of these practices, which increase costs, to competing industries reduces one of the constraints on the union imposing them. The rationale is identical to the explanation of the failure of featherbedding in pure competition. The market restraints on unions are diluted when competitive industries work under similar circumstances. For example, it is clearly advantageous to the railroads to have airlines face the problem of manning requirements on jet services, as well as the problem of restrictions on abandonment by public authorities. A great comfort to the Brotherhood of Locomotive Firemen and Helpers should be the increase in the crew consist on some commercial airlines resulting from the joint demands of flight engineers and the Airline Pilots Association.

The unions involved in featherbedding are narrowly organized along craft lines. There is almost a total absence of these rules in industrial or multicraft unions. The economics of this is important in understanding the role of the market in eliminating restrictive practices and restricting their introduction. The advantages of featherbedding accrue to a specific group, and this serves as a restraint upon gains to other workers in the firm. In an industrial union small groups rarely elicit the support of the entire organization for their own narrow ends, because the larger group has nothing to gain.

Union rivalry also leads to featherbedding as a defensive strategy. Featherbedding is found when there is a cluster of craft unions in an industry. These unions abide with each other under unstable conditions for a number of reasons. The organizations are competitive, pursuing

<sup>1</sup> Norman J. Simler, "The Economics of Featherbedding," *Ind. and Labor Rel. Rev.*, Oct., 1962, p. 100.

individual goals with little regard for the impact of their policies on nonmember employees.

A current dispute in the airline industry concerns the manning of the third seat in the cockpit. Should it be a member of the Flight Engineers with some pilot training or a pilot with an engineer's license? One result of course has been the three pilot-one engineer crew on some lines. The Emergency Boards examining this dispute have pointed to the need for union merger as a necessary condition for the settlement of the problem.

While employment goals of unions are not frequently dominating, much of the literature on trade-union utility functions implies that this is an aberration. These older models of trade unions, like those of Fellner, Ross, and Dunlop, were influenced by the flush labor markets of the 1940's. Such a position can hardly be accepted in the light of recent labor experience. While an interest in employment has long been noted in craft unions, one notes the growing interest of industrial unions in employment security. More and more we observe union leaders placing increased emphasis on employment factors rather than on wages. The type of activity pursued by unions with employment interests varies markedly from featherbedding at one extreme to the program recently effected by Kaiser, the ILA-PMA and Armour. It would appear that there is a series of alternative trade union utility functions and that there is a need for a systematic analysis of why a trade union chooses one alternative rather than another.

The technological requirements for featherbedding are at least as important as the industry and union structure. An almost universal characteristic is the gradual substitution of one form of technology for another. The displacement is evolutionary, wiping out the wage-rent differential of skilled groups. This type of change establishes the conditions for the imposition of the rules, as well as pointing the way toward their elimination. Thus, the type of rule we are concerned with arises shortly after the commercial introduction of a technology that is likely to adversely affect a relatively small and usually skilled, group in the work force. Quite frequently featherbedding emerges from the carrying forward of a set of practices appropriate for one technology to another where it is alien.

The specific labor groups engaged in featherbedding do not allow employers to modify the job assignments of the workers. This fact along with analysis of broader categories of inputs implies that the elasticity of substitution for specific groups in a firm is zero or close to it.<sup>2</sup> To

<sup>2</sup> K. J. Arrow, H. B. Chenery, B. S. Minhas, and R. M. Solow, "Capital Labor Substitution and Economic Efficiency," *Rev. of Econ. and Statis.*, Aug., 1961, pp. 225-50; also, Jora R. Minasian, "Elasticity of Substitution and Constant Output Demand Curves for Labor," *J.P.E.*, June, 1961, pp. 261-70.

analyze the effects of featherbedding with the aid of a Cobb-Douglas function<sup>3</sup> having a positive and constant elasticity of substitution for other factors appears most inappropriate. The results using that model imply that the marginal productivity of capital increases as a result of featherbedding—a conclusion that evaporates when a more realistic production function is employed. However, this is an empirical problem, and one that needs examination before a definitive answer to the impact of featherbedding can be provided.

## II. *The Impact of Restrictive Rules*

There are two avenues of analysis open: static analysis, which is fairly well developed, and dynamics, which is largely in an embryonic state. Some conclusions about featherbedding under static conditions are summarized, while some interesting dynamic questions are posed.

Theoretically a featherbedding rule could require a fixed amount of labor to be hired for a specified economic period. Under this procedure labor would be a semivariable cost and thus be analogous to any other lumpy factor.

However, the rules considered under the rubric of featherbedding do not specify that a fixed amount of labor be hired. On the contrary, they usually call for the retention of an existing labor-output ratio. Examples are the "bogus" role, double heading and standby. In each case the amount of redundant work to be performed by the firm is a function of output, thus affecting the height and slope of variable costs and, given the structure of the industry, reduced output. Consequently, we can say that the rules do reduce output and raise costs and prices in a short run.

The first dynamic question concerns the impact of featherbedding on supply. Does the existence of featherbedding affect the labor supply function? Unfortunately this is not easily answered, but let me point out the implications under either of two sets of conditions. First, if we assume that there is some onus connected with a job that is reputed to involve featherbedding, then the supply to the industry would shift to the left and in the extreme would possibly disappear; i.e., would be perfectly inelastic at zero offerings over the relevant wage range. In the long run, the problem might then be self-correcting from the supply side. An alternative would be for the wage rate to be adjusted upward to compensate for the disutility of being in an undesirable occupation. Thus, a consequence of featherbedding would be an upward pressure on wages that cannot be stopped in a short run. If the rule does not adversely affect supply, then the problem can never be self-adjusting

<sup>3</sup> Norman J. Simler, *op. cit.*, pp. 96, 97.

from the supply side. This obviously opens the question to some other type of policy.

A second dynamic problem concerns the impact of the rules on the rate of technological change and investment. It is commonly assumed that these restrictive rules retard progress. Clearly the intent of the rules is to lower the marginal efficiency of investment on laborsaving changes. The employer is not free to reap the full cost-reducing advantages of the change and unless the demand for the final product is perfectly inelastic some of the increase in cost must be shifted to ownership or other factors. The meager work on this problem is inconclusive, but indicates that the rules have some, though marginal, negative effect. I would argue that it is theoretically possible for the rules to stimulate change under certain conditions.

Let us assume that there is a new technology that allows the production of a product without the use of a factor used in the old technology. Assume further that there is a rule requiring a fixed labor-output ratio for one of the factors. Then the level of the featherbedding cost as measured by the slack variable depends on the mix of the two techniques. In the early stages of the substitution of the technologies there is little of the new and much of the old. Therefore the cost of the rule is minor. However, the more the new technology is substituted for the old, the higher is the relevant featherbedding cost. The maximum cost is obtained prior to the complete substitution of the new technology. When the last unit of capital using the old inputs is scrapped, the workers are entirely superfluous. The absolute cost of featherbedding through time depends on the length of time it takes to complete the change in technology. If this is perceived, then the featherbed rule should foster the more rapid introduction of a new technology and one that is radically different.<sup>4</sup> While research on this is not complete, it is suggested that this model is fruitful in explaining the history of the dispute concerning crew consist rules in the railroads. The critical factors are thus the character of the new production function and the time period between introduction and total substitution.

A less optimistic result comes about if it is assumed that some skills are still required even under the new technology. In this instance one sees that the problem is not corrected from the demand side and in fact may grow at a rate consistent with the increase in some parameter, such as the amount of matrix work in the printing trades.

In the former case, that involving the complete change in technology, the market can resolve the problem of the firm in the long run, but not in the latter case. Given that there are supply and demand conditions

<sup>4</sup> This assumes absence of cooperation by other unions. *Supra*.

which may forestall a resolution of the problem, is there any way out of the featherbedding dilemma? One might also point out the fact that even though the problem say from the supply side is soluble in the long run, it may be deemed that the period is too long, both from the firm's position as well as in tying up manpower resources.

### III. *Private Programs*

The threat of technological unemployment can give rise to programs to ease the adjustments. Recent experiments such as the Armour, West Coast, and Kaiser agreements have been proposed as models to be followed in other agreements. The advantages of these plans are that they are arrived at privately and do not countenance redundant labor. It would appear that the market is effective in ending the problem of featherbedding by providing a more attractive alternative to the parties. Are these programs desirable alternatives? If they are desirable, what conditions are necessary for their adoption?

Longshoring has a fragile existence as a separate industry or trade. The rationale for a separate union stems from the historic irregularity of shipping, yielding a fluctuating labor demand and with low integration in the industry, separate and corrupt unions.<sup>5</sup> Whenever the trade is regularized, as in coastwise shipping, or the market decasualized, through restricted entry in the hiring hall, the unions cease to have a separate identity and become submerged into teamster or ocean shipping unions. The restrictive rules have arisen out of the corrupt, casual labor market and have tended to disappear with organization. Faced with technological change, as well as hostility from the Teamsters, the ILWU leadership pressed demands on their membership for a program that allowed management to institute work rules changes. The *quid pro quo* was a management financed fund of some \$29 million to be used for early and regular retirement benefits, death benefits, and the stabilization of workers' income against declines due to changes in technology. The primary burden of disemployment resulting from technological change was to be borne by the workers last to arrive in the industry, many of whom are not members of the union. The union agreed not to object to alterations in rules and new technology, except in the case of speed-up. However, the ILWU have opposed changes quite regularly, as shown by the heavy use of arbitration.

The agreement has been advantageous to the parties. Management has added flexibility. The union has ended the hostile period with management and can concentrate on protecting itself against the Teamsters and resolving international problems.

<sup>5</sup>The corruption arises from the ease of discriminating pricing.

In the meat industry, the Armour program has devoted considerable energies in studying displacement problems caused by plant shut-downs or elimination of divisions, as well as alternative ways of dealing with the workers' problems. The Armour Automation Committee has sponsored useful research that has resulted in some steps that have aided the displaced. For example, an early warning system for plant shut-down and the T.A.P. program. The activities that have received most attention are in the labor market sphere of retaining and placement.

Featherbedding was never an alternative open to the meat industry unions. The structure of the industry precluded such strong union policy. The emergence of new plants and firms in geographic areas that are difficult and costly to organize, as well as lower entry barriers, has made the industry more competitive.

The unions in the meat industry show increasing dissatisfaction with this endeavor. This plan has been in jeopardy because of union frustration that reflects their inability to enforce an employment guarantee solution. It is also important to note that this is the only plan that has actually lived with the problem of structural unemployment.

The Kaiser sharing plan has as one of its activities a program for employment security in addition to productivity sharing. The plan allows the employer to introduce new techniques and alter work rules, but alleviates the insecurity usually associated with this. Workers displaced by technological changes are bumped into the plant-wide pool, maintaining their former wages.

There has been no real test of the plan, and the high attrition rate at the Fontana plant makes it unlikely that the pool will be too costly. The high attrition rate, estimated at 8 percent, and the twenty-six week eligibility requirement were originally viewed as being in excess of manpower reductions due to altered operations. The company has bought the freedom to adjust procedures with an employment guarantee of dubious value. The union has acceded to changes in technology elsewhere and in fact is not known either for featherbedding or particular concern about permanently unemployed steelworkers.

The reason for this lack of interest, in addition to the character of the union utility function, is basic to understanding the limited role of private agreements in handling the displaced. These programs are adequate for avoiding conflicts but are not in fact perfect substitutes for featherbedding, as the institution of featherbedding or its continuance in the case of longshoring was not a real alternative.

The cause of featherbedding is employment insecurity. The restrictive rules maintain jobs for the displaced. The alternative for featherbedding is the placement of displaced workers in other jobs and this normally means outside the firm or industry. While the Armour pro-



gram attempts this, it is at once hampered by the inadequacies of public services and is an alternative to public services. Labor market functions performed by either the former employer or union are likely to be on a crisis basis. To adequately move workers from their old jobs implies careful planning based upon labor market surveys, employment projections, and training techniques in diverse industries. To invest in this activity would require a considerable incentive, and it is on this point that the future of these plans in handling employment insecurity breaks down.

The return to both labor and management of shifting workers smoothly out of their own province is either zero or close to it. What incentive is there for a particular union to invest in preparing workers for membership in another union and, a fortiori, management investing for the benefit of other firms. Today the Armour Automation Committee is resorting to public services more and more—for these are best suited for handling the real problems. The main contribution that the programs can make is in mobilizing the public service. However, a really efficient service would hardly need this catalyst.

#### IV. *Conclusion*

While the market through substitution can end the problem of redundant labor under specified conditions, it can only offer a superficial answer to the real problem of union insecurity. Both the limited interest of labor and management and the technical requirements of appropriate labor market policy require that a more vigorous program of public labor market activity be adopted. Further, it is to the interest of the economy to broaden the industrial basis of union organizations. The market does not seem to aid, and in fact may retard, broader trade-union organization. Without the broadening of their scope, even plans for labor market activity as positive alternatives to feather-bedding may collapse against the weight of organizational self-preservation.

## DISCUSSION

WILLIAM G. BOWEN: In accord with the division of labor suggested by our Chairman, I shall confine my comments to Professor Raimon's paper.

The policy observations presented in his paper rest on certain assertions about the operations of labor markets, and so it is perhaps best to begin by considering the most basic of these assertions; namely, that "anticompetitive institutional pressures" on wages of various groups (especially production workers in manufacturing) seriously impair the effectiveness of relative wages as an allocative mechanism. Let me say at the outset that, in company with most of my fellow economists, I accept the general thrust of this argument. What is difficult for me is to arrive at judgments with regard to the magnitude of this problem and the policy implications.

Raimon has not pretended to present the results of new, intensive research which would ease these judgment problems. Instead, he has cited earlier labor market studies as offering "strong support for this anticompetitive institutional control of the labor market." His interpretation of this literature seems a bit one sided, however, in that he does not mention the studies of Rees and others, which show the importance of underlying market forces and emphasize the ease with which one can exaggerate the role played by institutions as such.

Also, Raimon's assertion that there have been "repeated observations of an absence of short-run positive correlation between wage and employment changes by industry" seems to me to be somewhat misleading. The Eckstein-Wilson study makes use of an industry-unemployment variable as one of two main explanatory factors; and my own empirical study of postwar wage behavior revealed evidence of a positive correlation between industry employment and average hourly earnings of production workers during relatively good times (but not during periods of high unemployment). Finally, it is worth recalling that most of these empirical studies stress the intercorrelation between profits and employment, and the resultant difficulty in identifying the role of employment changes alone.

In addition to mentioning past studies, Raimon supports his inefficiency-of-the-relative-wage-mechanism argument by citing figures showing that between 1947 and 1961 disbursements per salaried worker in manufacturing increased at a slower rate than comparable earnings for production workers, even though salaried worker employment was expanding and production worker employment declining. In interpreting figures of this kind, it should, I think, be borne in mind that part of the explanation may well run in terms of the relatively high elasticity of supply of female salaried employees, who are more apt than men to move easily into and out of the labor force (rather than into and out of unemployment) in response to changing labor market conditions. Furthermore, since World War II there have been a number of factors operating to produce a pronounced rightward shift in the supply curve for female labor.

By these observations I do not mean to imply that I disagree with Raimon's basic point. Both previously published research and the more recent figures he cites suggest that relative wages are less sensitive to balances between supply and demand than would be desirable from the resource allocation standpoint. I do mean to suggest, however, that I think Raimon overstates the point, that relative wages are not as unresponsive as he implies.

I also have the impression that Raimon overstates the role of unions in impeding wage flexibility. Many studies have suggested that the policies of large companies (generally in the manufacturing sector) are themselves a source of wage rigidity. Such companies are often anxious to preserve their reputation as "good places to work," are cognizant of the investment they have made in trained workers, and are conscious of the cost of changing complex wage structures in response to what may be short-run changes in demand and supply conditions.

Looking ahead, Raimon sees no reason to expect collective bargaining to produce more flexibility of the wage structure type. Be this as it may, collective bargaining is covering a smaller and smaller proportion of the work force each year, and so collective bargaining as an institution may be expected to become somewhat less important as a determinant of the overall wage structure. Also, as (if) the overall level of unemployment falls, I expect that we shall see more relative wage changes in the "right" direction than occur when total unemployment is high. This guess is based on the premise that as general labor market conditions improve competition for skilled workers and white-collar employees will push up their wages relative to the wages of production workers faster than during periods when total unemployment is high.

Turning now to policy issues, we should first note that Raimon is concerned about relative wages and labor mobility primarily from the standpoint of the unemployment-inflation problem, not from the standpoint of efficient resource allocation in the Pareto sense. He asserts that our country's tolerance for unemployment is lessening, and right here I must register a dissent. It seems to me quite probable that our prolonged exposure to 5+ percent unemployment has led to an upward revision of what most people consider to be "normal" levels of unemployment.

I agree with Raimon that there are no obvious, practical ways of significantly increasing relative wage flexibility, though, as stated earlier, I am more optimistic than he is as to the outlook for wage flexibility, even given no changes in present institutions.

I am most seriously bothered, however, by Raimon's closing policy assertion: that countering rigidities and immobility by training programs which increase the productivity and employment prospects of individual workers need involve no inefficiencies. To be against retraining these days is like being against love and good works, and I am not. But I do suspect that some of us, most of whom are in the business of teaching and training, have a natural tendency to exaggerate a bit the likely success of massive retraining schemes. Training is not, of course, costless; and I am convinced that in this area as well as all other areas one can reach a point at which the rate of re-

turn is sufficiently low to make further expansion of the activity uneconomic. When such a point is reached (and I am not suggesting we are there now), to continue to attack an allocation problem by attempts at retraining can be grossly inefficient.

GEORGE H. HILDEBRAND: Professor Weinstein has sought to provide a theoretical analysis of featherbedding: the conditions under which it appears, its impacts, and alternatives for dealing with the underlying problem of labor displacement. I am in agreement with much that he has to say, and applaud his courage in facing up to the issues instead of retreating into a lot of neo-Luddite sentimentality.

However, I find his approach unsatisfactory in some respects. Fruitful study of the featherbedding question requires, first, that the problem itself be clearly identified and, second, that attempts to formulate a "pure theory" of featherbedding ought to follow rather than to precede a detailed study of the evidence. The main drift of economic theory today too frequently departs from this simple canon of good science. The sin of the institutionalists was that they put too much emphasis upon gathering the facts first, although they did comparatively little in this regard, and that they had too little theory with which to inform and guide their efforts. The opposite fault dominates the scene today: too much theorizing *in vacuo* and not enough testing of implications. I do not suggest that Professor Weinstein is a full-fledged sinner in this respect, but I do think that his approach would be strengthened if he had put off generalizing in favor of more detailed study of the phenomenon with which he is concerned.

I cannot find in his paper an explicit definition of what featherbedding is. Is it to be identified with continued enforced manning of obsolete jobs, with excessive crew sizes, with limitations upon output, with obstruction of new techniques, with required performance of unneeded work, with strict partitioning of tasks and work jurisdictions, or with some or all of these together? These practices vary from industry to industry. Not all of them take their rationale from make-work objectives. Often they represent compromises over conflicting equities regarding fair work loads. Absent standards of absolute justice in these matters, it becomes very difficult to isolate and to measure the make-work component. Beyond this, even where make-work is the guiding intent, its origins can be diverse, and no single, unified theoretical explanation will account for all cases.

To illustrate, Weinstein attributes featherbedding to "employment insecurity under special circumstances." One of these, he says, is that the industry has to be "noncompetitive" before the rule is introduced. Another requirement, he asserts, is that there be present a multiple structure of narrow craft unions. Another is that a new technology gradually must be displacing an older one.

Exceptions exist here all along the line. The railroads were by no means a monopolistic industry in 1939 when the freight diesel came into over-the-road use. They are not so today. But they do have a problem of unneeded firemen, because federal and state legislation has denied them all opportu-

nity to solve it. Take another case: industrial unions in manufacturing in some instances have prevented employers from contracting out maintenance work. Are we to conclude that these industries are noncompetitive and have a craft structure?

The origins of restrictive rules are also diverse and need not derive from shrinking job opportunities, granted that in some cases they have. On the railroads, the fireman's job is almost wholly obsolete. But given the numerous shock-absorbers the roads have offered these men to ease the transition, the problem is not basically one of job and income protection. Rather, it is a political one. The leaders of the union simply do not propose to go out of business except for passenger runs.

Further insights would be gained, I believe, from sorting out more carefully the underlying causes of employment insecurity. Does it arise from technology alone, as with the fireman? Is it made worse by a secularly declining product market, as is now affecting all employment on the railroads? Or is it the consequence primarily of anticompetitive wage raising, which creates labor surpluses and leads to work rationing and restrictive devices, as in the construction trades? Obviously, the causes of insecurity can be diverse. They issue in different types of featherbedding, and they call for divergent solutions to the problem itself.

I turn now to the impacts of featherbedding. Weinstein argues in the abstract that restrictive rules may accelerate the substitution of a new technology. His case is persuasive. As the changeover proceeds, the cost of redundant labor rises, reaching its peak just before transition is complete. Thus the employer has incentive to get rid of the loss as soon as he can. To Weinstein, these principles suggest a fruitful way to approach the problem of "crew consists" on the rails, concluding that "the market can resolve the problem of the firm in the long run." But matters will not turn out so happily, he adds, if the new technology retains a requirement for some of the old skills.

A priori, rising costs ought to lead to efforts to economize, here to get rid of unneeded hands. But this market force is strangely absent on the railroads, and for good reason. Since 1956 the carriers have been trying to eliminate the freight fireman, to introduce a realistic definition of a fair day's work, and to obtain other modernizing reforms in the use of manpower. They initiated their belated efforts at that time because their net revenue had begun to slip alarmingly. Control of costs had become the price of survival, and so a frontal attack upon now intolerably wasteful ancient practices became mandatory. Even before this time, the roads had achieved a remarkably rapid increase in labor productivity, one that has been well sustained. However, practically all of it has been won by indirect devices: bigger and faster power, longer trains, larger equipment, less helper service, more efficient sorting yards, highly mechanized off-track maintenance, and ruthless elimination of extramarginal services. In the outcome, these changes have yielded more ton-miles and more revenue per labor hour worked.

But virtually none of them has involved revisions of job structure, of manning requirements, or of the amount of time to be worked in a basic day. There are two reasons for this. First, the unions have steadily refused to take a

long-run view, to admit that total job opportunities depend above all else upon the ability of their industry to compete against heavily subsidized, undertaxed, and largely unregulated competition; that is, upon its ability to cut costs and rates, and so enlarge its share of the product market. In this respect, featherbedding has hastened the attrition of employment by the grim process of speeding the demise of the industry itself—hardly the long-run solution contemplated in Weinstein's model. Undoubtedly, as he suggests, this blindness of vision in the leadership of the railroad unions derives in some part from the craft structure with which the industry is beset. There also seems to be a price that must be paid simply for being old.

Second, the firms in the railroad industry have been barred from a long-run solution by detailed and extreme statutory control of their relations with the unions. For years there has been no real collective bargaining in the industry because public policy rules out the strike, and because the whole system of intervention generally has favored the unions' cause. The recent case of the Florida East Coast is a perfect illustration of the point. And so we have the tragic spectacle of a great industry, replete with high promise of efficiency superior to its competitors in so many ways—as ample foreign experience shows—undergoing slow economic strangulation at the hands of the unions and the government. The work-rules case is now approaching its ninth year. When will the noose be loosened, if before the point of death?

Let me turn now to Weinstein's views about the West Coast longshore plan. I disagree that the longshore attrition scheme does not countenance redundant labor. It requires that all those on the present registered list continue to be offered employment and to be guaranteed standard weekly earnings, regardless of required time to be worked under the more liberal work rules now coming into force. Attrition will eventually wipe out this excess cost, but as of now it must be borne. It arises from the continued use of unneeded men.

Also, I do not understand the assertion that with hiring halls and decasualization the longshore locals became "submerged into teamster or ocean shipping unions." I do not know whether this statement is intended to be descriptive or a characterization. However, it is a fact that the International Longshoremen's and Warehousemen's Union started life in 1933 as a branch of the old ILA, becoming independent within a year or so, and remaining such today. It is not a part of the Teamsters, nor of any of the seafaring crafts. Furthermore, the statement that in longshoring "the restrictive rules have arisen out of the corrupt, casual labor market and have tended to disappear with organization" simply does not mean what it says. One of the basic issues in the great strike of 1934 was speed-up under nonunion conditions. From the settlement of that dispute until 1959, the union fought mightily and successfully to regulate gang sizes and make-ups; sling loads; sorting, piling, and stowage; and the boundaries of its work jurisdiction. These rules restrained the employers' freedom of action, and one of their clear purposes was to conserve the work opportunity. It took a quarter-century for the union to change its position.

Beyond question, the ILWU's new approach was a courageous and forward-looking step, introduced into an industry where, as with the railroads,

employment had been slowly shrinking for some years. Indeed, this was the very reason why the leadership decided upon a change of stance toward efficiency. On the railroads, I note for the record that the carriers accepted the same basic principles as are embodied in the West Coast plan, as put forward by the President's Commission early in 1962, as made somewhat more generous toward the men by the Emergency Board in 1963, and as liberalized further by the special tribunal created by statute last August. In short—and not without misgivings—the carriers offered to use a substantial part of their prospective savings from modernization of work rules to ease the shock of transition. So far, they have gotten nowhere, partly because they are enmeshed in interminable legal proceedings, and partly because, as Weinstein suggests, the craft structure of the unions fosters obdurate resistance. If the roads are not to be allowed to follow the strategy of Eastern Air Lines toward the Flight Engineers, then some kind of statutory solution to the problem of obsolete bargaining structures seems the only alternative.

Weinstein offers the concluding generalization that "restrictive rules maintain jobs for the displaced." This statement requires some qualification. These rules keep average variable costs at too high a level for viable competition. In a keenly competitive field such as railroad transportation, the effect is not to preserve jobs, but to destroy them. The reason is that noncompetitive costs compel steady attrition of operations. In truth, therefore, the only ones who get any protection are the survivors, who could survive without it. The losers are the younger men, who are forced out as the very by-product of keeping the rules in force.

However, I would agree that we require much stronger public programs to improve the functioning of labor markets and to aid the mobility of labor. Wage and work-jurisdiction policies being what they are, the burden laid upon mobility is made far heavier than it would have to be under full freedom of competition. But the basic problem is not one of greater mobility alone. Our economy is beset with a deficiency in overall demand for products as well as for labor. Structural solutions will not solve the employment problem by themselves. More than this, they will work more successfully if the trend rate of growth is increased.

I turn now to a few remarks about Professor Lewis' highly important effort to estimate the relative wage and employment effects of unionism. Unhappily, I must confess to some bafflement in attempting to understand his paper, because the underlying argument is unclear and obscure at several points. In part, no doubt, the reason is that the method employed to obtain the basic interindustry series in Table 1 is not actually set forth. Instead, the reader is referred to the author's new book, published this very month, and which I have not been able to see. In consequence, most of my comments will take the form of questions. Perhaps these are also inspired by my own limitations in mathematics.

Lewis is concerned with the impacts of unionism upon relative wages and relative employment. He attacks the problem first on an aggregative level, by comparing two sectors. In Group *a*, the percentage of unionization has been high throughout the period considered, while in Group *b* the contrary has been

true. My first question is, why were salary employees included in both sectors? Was it for limitations of data? Given the lower propensity for these workers to join unions, their inclusion would seem to weaken the validity of the findings—how much, I do not know.

Next, columns 5 and 6 refer to comparative extent of union membership. The method of computation is not clear to me, nor is the underlying evidence set forth. Do I assume correctly here that a value of unity in some given year would mean one hundred percent unionization in Group *a*, and zero unionization in *b*? Again, I do not know.

I am also unclear about the meaning of columns 7 and 8, the estimated relative wage effects of unionism over the years. Lewis says that the effect is expressed in common logs "per percentage point difference" in comparative unionization, presumably of the two sectors. If so, the reference may be to a regression coefficient per point of difference, which would be an estimated constant, I should suppose. But since the quantities in columns 7 and 8 are varying over time, other factors besides percentage point differences may be involved. Would one of these be the actual decimal point differences in unionization, year by year?

I come to a more important problem. What is the basic equation relating logs for relative wages to the variables upon which wages are said to depend? How were the unemployment and price-level variables estimated and treated? How good a fit was obtained for this relationship? And how is this missing equation connected to the labor demand and supply equations? Do we have, in fact, a simultaneous equation model in which all three are solved together, and their reciprocal interactions hence recognized? If we take the two demand and supply equations and rearrange terms slightly to get rid of the brackets, we are confronted with a puzzle. The coefficient *a* turns up in both, first as part of the term  $\log W_t$  on the demand side, and again as part of the term  $B_t P_t$  on the side of supply. Does this mean, as it suggests to me, that the relative wage effect on labor demand is the same as the relative unionization effect on labor supply? There must be an underlying theory to this, but I do not find it here.

However, the main conclusions following from the aggregative study are not ones with which I would quarrel. I think it correct that unionism has raised relative wages and has lowered relative employment. But I wish that the exposition were fuller and more understandable.

I have a few words to say regarding Lewis' findings for the bituminous coal industry. To estimate the employment effects of unionism here, he splits them into three components. One of these is the substitution effect. He adopts the assumption that constant elasticity of substitution  $\sigma$  prevails. Presumably this was in the form developed by Arrow, Chenery, Minhas, and Solow (*Rev. of Econ. and Statis.*, Aug., 1961). Are we then to assume, as I understand these authors' formulation, that in bituminous coal perfect competition, equilibrium, and constant returns to scale continuously prevailed in this period? If so, that is a large assumption indeed.



# THE REGULATED INDUSTRIES

## THE EFFECT OF RATE REGULATION ON RESOURCE ALLOCATION IN TRANSPORTATION

By GEORGE WILSON  
*Indiana University*

The consequences of rate regulation depend upon the type of such regulation, its effectiveness, and extent. Since we are worried about misallocation within transportation, which implies a concern with the distribution of traffic among the various competing modes, I shall concentrate on one regulatory body, namely, the ICC, since it has jurisdiction over those modes of transport deemed to be most closely competitive—in particular, rail, truck, and, in some instances, water. Furthermore, the most important and analytically interesting questions are those involved with rate policy as it affects specific commodities between particular points. I will therefore resist the temptation to make caustic comments about the regulation of what is loosely referred to as the general rate level.<sup>1</sup>

It is not easy to determine what ICC rate policy in fact is. Although exceptions can be found, the following appear to be the most frequently used criteria. I will examine each to ascertain the extent to which regulation, even if complete, creates differences from what might be expected in the absence of regulation and then assess the allocative implications.

1. *Each Case Sui Generis*. Although from time to time the Commission may enunciate certain general rules, these are normally put in a *ceteris paribus* context. Since transportation markets, properly construed in terms of city-pairs, are characterized by substantial diversity, there are in fact few general principles that are directly relevant to particular circumstances. But in this sense, the situation is precisely what it would be in the absence of regulation. If firms are rational, they

<sup>1</sup> Few persons are concerned any more over railroads making "excessive returns." There is more interest in truck profitability, but the Commission artfully covers this up by using the operating ratio as an index of profitability. Comparison with railroads is further confused by using return on investment as the index of rail profitability. If, however, we take the ICC's conception of normal profits in trucking to be an operating ratio of .93 and normal rail profits to be 6.5 percent return on investment, the interesting result emerges, on certain normative assumptions regarding capital turnover, that equilibrium ensues when rails earn 6.5 percent and trucks earn over 30 percent, a conclusion fraught with intriguing implications which cannot be pursued further here. It seems to me, however, that this issue is of some practical as well as theoretical importance. Indeed, the origin of the problem is to be found in Adam Smith where profits are construed in both senses.

will adjust rates to meet the exigencies of particular times and places. Each rate change, even when based upon profit-maximization principles, will differ because the marginal costs and revenues relevant to a particular transportation firm depend upon the demand for and nature of the commodity to be shipped, the rates and service of competing forms of transport, particular shipper and carrier circumstances, and so on. Even under regulation all of these considerations enter into any pricing decision, although they are frequently accorded different weights by the Commission than would be the case without regulation.

2. *Rate Floor at Out-of-Pocket Costs.* One of the major purposes of minimum rate regulation is to prevent rates from falling below something referred to as out-of-pocket costs on the assumption that rates below this level constitute an unfair or predatory method of competition. Out-of-pocket costs may appropriately be taken to refer to marginal costs;<sup>2</sup> hence, rate regulation seeks to prevent any rate lower than marginal cost.

The Commission likewise makes a distinction between short-run (i.e., immediately escapable) and long-run marginal costs. Although there are serious analytical, interpretive, and statistical complications regarding this distinction, it is clear that the Commission construes the rate floor in terms of at least a proxy for the economist's concept of long-run marginal cost and that it frowns upon any cost analysis that includes only those costs that are immediately escapable. Without rate regulation would there be a tendency for rates to fall below long-run marginal costs? The answer to this depends upon two main factors: (1) knowledge of specific costs, including cost behavior, and (2) the long-run profitability of predatory pricing.

Our knowledge of specific costs and cost behavior in transport is most inadequate. Despite important advances in recent years, none of the techniques of cost finding hitherto employed yields results that have any necessary connection whatsoever with what the economist means by long-run marginal cost. Not only do all the empirical studies use accounting cost data which may deviate substantially from economic costs, but for rail, barge, and pipeline transport in particular

<sup>2</sup>There have been, from time to time, doubts raised concerning the equation between out-of-pocket costs in the Commission's sense and marginal costs in the economic sense. (G. L. Wilson and J. R. Rose, "Out-of-Pocket Cost in Railroad Freight Rates," *Q.J.E.*, Aug., 1946, pp. 546-60. More recently, M. J. Roberts, "The Regulation of Transport Price Competition," in *Law and Contemporary Problems, Transportation: Part I*, Autumn, 1959, p. 568, f.n. 39, has restated the argument of Wilson and Rose.) However, the Commission's definition is essentially the same as the appropriate economic conception; namely, the increment in costs associated with the increment in output. The basic difficulty, like so much else in transportation economics, resides in defining a unit of output as distinguished from a unit of sales. These and other points are elaborated more fully in my *Essays on Some Unsettled Questions in the Economics of Transportation* (Foundation for Economic and Business Studies, Indiana University, 1962).

the significant gap between the sales and output unit creates problems of cost indivisibility over and above those other cost categories traditionally deemed nonassignable to specific output (and especially sales) units on a cost-occasioned basis; namely, fixed and joint costs. It is frequently forgotten that all variable output costs are divisible or traceable to specific sales units only if the output and sales units coincide. Where they do not, many variable output costs become indivisible. Unlike the problem of cost fixity, the magnitude of both the indivisible and joint costs is independent of time. Although the appropriate conception of what actual, calendar time span constitutes "long run" remains a serious difficulty, especially in empirical cost studies, it is clear that the relevance of time from the point of view of cost assignability is in deciding what amount of output costs are variable. If there exists a constant gap between pricing (or sales) units and output units, the variability of the number of output units does not affect the relative magnitude of joint and indivisible costs. In short, the long-run marginal cost of an extra sales unit (LRMC's) will regularly be below the long-run marginal cost of an extra output unit.

But even finding LRMC's requires a host of assumptions and choices prior to any statistical analysis. From an empirical point of view one must select the sample carriers in any cross-section analysis, the years for which the accounting data are to be used, the appropriate form of the regression function, the variables to use in determining coefficients, the degree of cost refinement, and so on. When all of these choices have been made, even if each is perfectly reasonable, the result gives a "sort of central tendency" when in fact what is required for pricing purposes is a set of estimates that encompasses the wide variety and variability of carrier and shipper circumstances for particular shipments between specified points. If there were less variety and change in this regard than in fact appears to exist, central tendency results would be good enough; but contemporary transport markets appear to contain substantial variety on both the cost and demand side and it is this variety as well as variability that ought not to be averaged away and that needs to be exploited.

For these and many other reasons it is not surprising to find many transportation economists as well as company officials lamenting the lack of good cost data that would provide appropriate guidance to pricing policy. But the point here is that costs in transport, as elsewhere, are not the objective, tangible things so frequently presumed. The avoidable economic costs in any particular circumstance depend upon a host of subjective factors, as well as technical and pecuniary facts. Cost, then, is a pretty elusive and ambiguous concept. This implies that

any management can demonstrate statistically that whatever rate it sets, or proposes to set, is above its own subjectively estimated long-run marginal costs. Thus, whether through ignorance or intent, rates in specific instances can easily fall below long-run marginal costs in the economic sense. Even under regulation, however, this situation would exist. The only difference is that the probability of below-cost rates would be reduced since regulatory authorities generally employ cost data that probably overstate costs for a particular, efficient carrier and for particular kinds of traffic.

However, the regulatory process, as distinct from its use of broad cost averages, would discourage below-cost rates even if cost is not adequately known. Respondents will generally determine the costs to be higher than the carriers initiating a rate decrease and in the ensuing compromise of statistical techniques and estimates a rate higher than originally proposed will often be ordered. Generally, the need to defend cost estimates tends to make these or the proposed rates higher than would be the case if a carrier merely sought additional traffic. At the same time, however, there is a strong incentive for the initiating carrier deliberately to employ techniques of cost finding and assumptions which yield low cost estimates relative to the rates and vice versa for protesting carriers. On balance, it is difficult to assess the net regulatory impact upon rates, for like cost itself, much depends upon particular circumstances, such as the care with which the parties to the dispute prepare cost estimates, the projections of traffic volume anticipated at the proposed rate, and so on.<sup>3</sup> Nevertheless, it seems reasonable to conclude that rate regulation would deter below-cost pricing, even in the absence of good cost estimates, and would eliminate obviously predatory pricing. On the other hand, in the absence of rate regulation, past history suggests that many carriers have found predatory pricing sufficiently profitable to warrant its use from time to time either to eliminate or discipline a competitor. The net effect of this aspect of rate regulation is therefore to reduce the frequency and extent of below-cost rates that would result either from ignorance or intent.

3. *Rates to Contribute a "Fair Share" Above LRMC's.* The Commission is normally reluctant to permit rates as low as LRMC's and, indeed,

<sup>3</sup>It must be stressed that costs (variable as well as fixed) and rates are intimately interconnected so long as long-run cost linearity is not assumed; that is, since marginal cost depends upon volume and volume is some function of the rate, then marginal cost is likewise a function of the rate. This suggests that there may be some equivocation regarding "cost-based rate making" if this is meant to imply that the level of any rate is made dependent upon a prior calculation of cost. This also suggests that if the ICC is to retain minimum rate powers, it must make independent estimates of the effects of the rate on volume of traffic. It is, therefore, inconsistent to argue, as some have done, that the ICC retain the minimum rate power but give up any attempt to substitute its own estimates of demand elasticity.

some amount must be added to LRMC's if any profit on output units is to be achieved. However, the amount to be added is interpreted rather loosely in terms of no lower than "necessary" or not so low as to "unduly burden" other traffic. The precise meaning of these phrases is ambiguous, but there are two main implications. The notion that each rate must be sufficiently compensatory so as not to burden other traffic unduly leads to an emphasis upon fully allocated costs in the belief that if a particular rate is below fully allocated or average total costs (ATC) other rates must be sufficiently above ATC so that total revenues equal total costs including normal profit. The frequent allusions by the Commission to the fact that fixed costs are no less real than variable costs becomes a rationalization of ATC as a more appropriate rate minimum in the sense that such a rate on particular traffic requires no "excess" revenue contribution from other traffic which may be hard to get.

Now this addition of a predetermined prorata share of the fixed costs to computed marginal cost is what has drawn the scathing critique of professional economists who rightly repeat Jevons' famous "bygones" dictum. But the Commission is perfectly aware that this is arbitrary and contrary to sound economic principles. As Commissioner Webb recently put it: "The allocation of constant costs to specific traffic . . . is essentially arbitrary. . . . There is no doubt in my mind that such fully distributed cost constructions are bottomed on economic fallacy. . . . [But] this form of economic nonsense . . . may be entirely sound from a regulatory point of view . . . [because] regulation is designed to achieve a number of important objectives, the value of which cannot be determined on purely economic grounds."<sup>4</sup> Ignoring for the moment the noneconomic criteria, it appears that in a very crude way the Commission is groping for some potentially tangible or measurable standard to determine how much should be added to computed LRMC's. However rough or uneconomic the ATC standard may be does not deny the fact that something has to be added to LRMC's even on strictly economic grounds. Furthermore, the Commission, as the occasion warrants, prevents rates as low as ATC and under some circumstances permits rates below ATC but above LRMC's. This implies that the Commission may be attempting to seek for the regulated carriers a rate on each commodity that maximizes the net within the constraints of equity among shippers and regions, normal overall profits to carriers, and the other noneconomic objectives of regulation.<sup>5</sup>

<sup>4</sup>"Costs for Rate Making Purposes," Remarks of Charles A. Webb, Member, Interstate Commerce Commission, before the Annual Meeting of the National Accounting and Finance Council of American Trucking Associations, Washington, D.C., May 13, 1963, mimeographed, pp. 7-8.

<sup>5</sup>For example, Commissioner Webb in a recent case argued that "inflated out-of-pocket

If this is the case, then the situation under regulation is not basically different from what the carriers would seek if let alone, except that personal and place discrimination would logically receive no consideration, nor would the other features of national transportation policy such as the "needs" of the post office and national defense. There is, however, an important exception, even ignoring noneconomic criteria. If two or more regulated carriers are competing for the same traffic, the Commission will prevent the unraveling of the rate structure and will generally establish rates higher than would be the case without regulation. It will force a kind of compulsory cartel and for commodities that have a low elasticity of transport demand will generally maintain rate levels well above any measure of ATC. Of course, it is possible that without regulation the oligopoly solution might prevail, but it is doubtful whether it can be as successful in maintaining as high a rate as a Commission supported compulsory cartel arrangement jointly exploiting aggregate demand inelasticity.

The impact of completely effective rate regulation, insofar as the present criterion is concerned, is therefore to maintain rather higher rates where the aggregative elasticity of transport demand is low than would be the case without regulation, since demand elasticity will appear higher to the individual carriers. This is probably true even where the rational oligopoly solution would exist.

However, an important misapplication of such discriminatory pricing is that form of pricing which links the rate to the value of the commodity. All too frequently the Commission as well as the carriers appear to believe that the value of the transportation service, even when all modes are considered together, is closely related to the value of the commodity. In fact, of course, the value of the commodity is only one determinant of value of the service and in many cases not the most important one. The value of the transport service depends on the level and elasticity of the demand for the product and the ratio of the freight rate to the delivered price of the commodity. More precisely, the total elasticity of transport demand is simply the product of the elasticity of demand for the commodity and the ratio of the present freight rate to the delivered price of the commodity. From the point of

---

cost computations will prevent carriers in some cases from making rates which enhance net revenues to the maximum extent possible. Rates which barely exceed out-of-pocket costs may be condemned in some cases because they do not maximize net revenue and are apparently motivated by a desire to destroy competition, but on the other hand, rates at such a level may be the only means to obtain a modest contribution to fixed costs on the traffic involved." (Lumber—California and Oregon to California and Arizona, I & S Docket No. 6933, decided Aug. 10, 1959, p. 57.) Again, the Commission in another case approving proposed rate reductions stated that "the reduction in the level of rates in a competitive situation such as here has as its purpose the maximizing of net revenues through increased volume of traffic." (Paint and Related Articles—Official Territory, I & S Docket No. 7027, decided Aug. 27, 1959, mimeographed, p. 9.)

view of an individual mode of transport or single carrier, the rates and service quality of actual and potential competitors as well as their expected reactions are also significant determinants of value of the service. The value of the commodity is then but one among several factors essential to a pricing policy aimed at absorbing as much shipper surplus as possible for a particular carrier or carrier group. The important determinant of the value of the service—a term most appropriately defined from the economic point of view as the maximum profit rate on any commodity between a given city-pair—involves, in short, the demand function facing an individual carrier or group of carriers if they act in concert, as well as marginal cost. There is no point whatsoever in attempting to set rates on the basis of the level and elasticity of total transport demand if in fact the demand function facing the group or firm is lower and more elastic due to actual or potential competition. Yet, in many cases, this is precisely what has been done even where the possibility of nonregulated carriage exists. I suspect that it is this type of thing for which not only the Commission gets its lumps but also value of service as a pricing principle gets disparaged. Yet, in fact, this represents misapplication of the principle. There is a vital distinction between the principle itself and misuse thereof, and the fact of the latter should not be used as evidence of the inappropriateness of the former. Nevertheless, the Commission is prone to argue that high-value products in some sense “should” pay differentially higher freight rates and there is no question that this ethical judgment makes no economic sense in many cases.

4. *The Share-the-Traffic Criterion.* Where the aggregate elasticity of transport demand is low and where the feasibility of unregulated carriage is also low, the Commission is reluctant to permit either the elimination of one form of common carriage or such a reduction in rates that the amount of shipper surplus jointly absorbed by the common carriers is seriously depleted. In seeking to enhance or maintain the profitability of common carriage, the Commission will often prevent rate competition and establish a set of rate relationships such that the various modes each get some share of the total business available at rates that may well yield greater total profits to each mode than if one got all the business at a lower rate.

There is little doubt that this policy, which may in fact be on the wane and which involves very tricky estimates of shipper preferences and costs of private carriage, constitutes an important difference under regulation. In effect, the Commission is seeking to find the limit price for the common carriers involved above which private carriage would be invited and below which aggregate profits would be reduced. The

share-the-traffic criterion thus implies an extramarket restraint whose motivation is uniquely bound up with the preservation of common carrier profitability by joint common carrier exploitation of demand inelasticities up to the limit price.

5. *Rate Stability.* The entire regulatory process, including the foregoing policies as well as the thirty days' notice requirement and investigation and suspension procedures, inevitably leads to less frequent rate changes than would otherwise occur. Although designed for purposes other than prevention of "instability" per se, the necessary effect is a high degree of rate stability or rigidity, depending upon one's valuation.

6. *Attitude toward Various Forms of Discrimination and Other Noneconomic Factors.* According to the law and ICC interpretation thereof, commodity discrimination is eulogized, place discrimination tolerated under special circumstances, and personal discrimination outlawed. But in economics, discrimination is discrimination, and there is thus no valid economic reason for these diverse attitudes towards differential pricing. In fact, the apparent moral repugnance manifest in the views regarding personal discrimination frequently leads to a kind of reverse discrimination by limiting, for example, quantity discounts to carload lots although this, too, seems to be changing. If large and regular shipments assist the economical operation of a carrier, they should be encouraged and not restricted by a set of noneconomic considerations which in the final analysis cause not only additional economic waste but do not have the desired effect, since private carriage is a feasible option where substantial traffic on a regular basis is involved.

Without regulation these divergent attitudes toward various forms of discrimination would not occur. Furthermore, the many other noneconomic considerations and externalities would also be irrelevant so long as carrier managements rationally pursued their profit objectives. In reality, of course, the existence of noneconomic objectives makes any decision of the Commission "reasonable" on some grounds. It may well be of some value to the nation that these noneconomic objectives be pursued. But the carriers should not be judged on strictly economic grounds if required to perform noneconomic functions, and as economists we are rightly concerned that whatever the objectives, they be achieved by methods that occasion least cost.

*The Nature of Rate Regulation: A Grand Benthamite Design.* The foregoing sections may be summarized as follows. Rate regulation as practiced by the ICC, if completely effective, would maintain higher rates on specific traffic than would otherwise occur, would involve greater rigidity, tend to preserve a demand-oriented rate structure,



reinforce misapplication of value of service pricing but emphasize commodity discrimination as the most effective way of covering the nonassignable costs, frequently allow more than one mode to participate in a given piece of business regardless of relative costs, consider a wide variety of noneconomic objectives and externalities, treat each case as *sui generis*, more effectively prevent rates from falling below marginal cost, and would force a kind of public accountability and greater cost consciousness upon the common carriers. Insofar as one can impute rationality to contemporary regulation, the overall impression is that the Commission seeks a rate structure such that all non-traceable costs including those additional costs arising from pursuit of noneconomic objectives and the added duties of common carriage are recouped from shippers on the basis of willingness and ability to pay; that is to say, the markup over computed out-of-pocket costs would be determined by the relative level and elasticity of the demand for transport of particular commodities. Individual rates would be adjusted from the maximum profit level by consideration of external economies and diseconomies as far as regulated carriers are concerned and overall profit limitation is implied. Rate regulation, if completely effective, thus attempts, albeit in a crude and implicit way, to assess the benefits to the regulated carriers as a group and weigh these against whatever externalities appear to emerge in particular instances. It is true, of course, that, unlike Bentham, the ICC does not explicitly add up the net benefits to each shipper, carrier, and community involved or affected and then strike the balance, approving those rate changes for which the sum of the net benefits exceeds zero and disapproving all others. But the frequent references to injury or noninjury to some shippers, regions, defense interests, or other carriers due to proposed rate changes, suggests that an interpretation of rate policy in terms of Bentham's felicific calculus may not be so farfetched as it seems on the surface.

In fact, however, the Commission cannot follow this grand Benthamite design even if such were clearly envisioned. As is well known, the extent of effective rate regulation differs widely among modes, and less than 3 percent of the proposed rate changes by regulated common carriers are generally investigated. This does not mean that rate regulation is ineffective because many of the cases have wide ranging effects. But it does suggest that modifications need to be made in the conclusions regarding stability, rigidity, and inertia fostered by the regulatory process. Likewise, the large unregulated segments, especially in truck and water transport, put serious constraints on Commission policy. Where unregulated costs or rates are below those of common carriage,

the Commission is forced to permit lower rates (close to computed marginal costs) than it would normally sanction. This also reduces the ability to maintain a pricing structure oriented towards the total transport demand for particular traffic and prevents sharing the traffic among common carriers where rates fall below sustainable levels for all modes save one. It also jeopardizes the fulfillment of noneconomic goals through rate regulation and reduces the ability to accommodate the revenue needs of the regulated carriers to external diseconomies as well. Since competition from unregulated carriage differs widely from market to market, this explains the shifting weight given to the several criteria from case to case. Where unregulated competition is intense, the Commission is forced to consider costs more fully, give them more weight, and abandon the share-the-traffic criterion, defense needs, and so on. On the other hand, where unregulated competition is less vigorous or feasible, rate levels will be oriented more toward what the traffic will bear and externalities will receive greater weight.

In short, because of incomplete regulation the grand design cannot be fulfilled. It is this fact that lies at the heart of the Commission's attempts constantly to narrow the scope of exemptions and oppose their extension. But the Commission persists in attempts to realize the grand design, even with limited jurisdiction. As a result each case is *sui generis* not only in terms of costs and revenues but also in terms of the influence accorded the noneconomic considerations and externalities.

*How Does Rate Regulation Affect Allocation?* It is impossible to assess the amount of misallocation involved in the foregoing. Economic cost estimates are not sufficiently accurate to enable us to specify with any degree of precision the wastes of regulation. Furthermore, many things contribute significantly to misallocation in transportation. There is, for example, no way of determining whether truckers pay their fair share of the publicly-provided right of way. Various forms of transportation are differentially taxed and restricted by the several states. Union rules and regulations affect carriers in a wide variety of ways. Massive public investments of different sorts result in discriminations among modes. Many other factors condition resource allocation in transport, and to separate that portion of waste attributable to rate control is clearly beyond the scope of present knowledge. Indeed, it may even be the case that, for example, if some motor carriers are paying more than their fair share of highway costs, the prevention of rail rate reductions, in rail-motor competitive situations, might even induce a degree of relative use more closely related to the real economic costs of providing the service even though relative accounting

costs suggest the reverse. Apparent misallocations occasioned by rate regulation might partially offset misallocation due to other reasons.

Moreover, in the absence of rate regulation there is no assurance that particular rates would exceed real economic cost and, in fact, it seems probable that many rates would fall to below-cost levels given the subjective nature of costs. This would provide as much uneconomic incentive as far as relative use of the varying forms of transport, industrial location, and so on are concerned as the artificially higher rates upheld by regulation, although for other reasons it is preferable to err on the low than the high side since the division of labor is still limited by the extent of the market. Rate changes would be more frequent without regulation and since these would not necessarily be based upon costs, the uncertainty engendered may inhibit investment in locations where on strict economic grounds it would be warranted. While I believe there has been an overemphasis upon the virtues of rate stability, nonetheless many shippers are less concerned about particular rate levels than in knowing that their rivals are paying comparable rates. If we could be sure that the rate differentials reflected even roughly real economic cost differences, there would be no legitimate cause for complaint. But this is precisely what we cannot be certain of. Indeed, the absence of all rate regulation is liable to result in the kind of rate secrecy, rebates, and so forth that was partly responsible for regulation in the first place. And this has important consequences for allocation among all transport-using industries.

Let us not forget that there are certain advantages to society of regulation as well as competition, even though both may be highly imperfect. We are, of course, a long way from an optimal mix of the two, but competitive pressures are clearly not absent under contemporary circumstances. However, the real case against economic regulation of transport today is the administrative impossibility of complete and effective regulation, the disincentive to change once a regulatory pattern has been established, and the existing collection of numbers gratuitously referred to as the "rate structure." The incomplete regulation imposes an artificial restraint and higher costs upon those carriers subject to ICC jurisdiction, and this occasions a relative over-expansion of unregulated carriage, leaving regulated carriers in the role of residual participants in many cases. Even though it cannot be demonstrated that competition in transport is such that far better economic performance would ensue should rate controls be eliminated, it is still the case that incomplete regulation may be worse than none unless it adopts a much higher degree of flexibility in adapting to particular changing situations than hitherto shown. But delay is probably inherent in the regulatory process, and complete regulation of all forms

of transport is infeasible. Nor can it be doubted that the present rate structure is unduly cumbersome, complicated, and extremely wasteful independently of its failure to relate correctly to either cost or demand factors. When firms can exist by taking a share of the gains to a shipper from a third audit of rate bills, when many shippers must maintain large staffs merely to assess the accuracy and legality of the rates actually charged, there is, to put it mildly, great need for tariff simplification which is unlikely to come from present regulatory practice.

It is mainly for these reasons that President Kennedy's transportation message makes a good deal of sense. Although the economic consequences of substantial reductions in regulatory restraint cannot be adequately nor unambiguously deduced in advance in the present state of knowledge, we would learn much from the experiment. But we should also be prepared to abandon the public utility aspects of transportation and the noneconomic features as well. This would be less serious if the experiment could be expected to result in improved efficiency and, indeed, the satisfaction of the noneconomic goals of regulation might better be attempted by other techniques, such as direct subsidy in the case of unprofitable services deemed desirable. Nevertheless, any industry having such widespread ramifications as transport and which occasions such enormous external economies and diseconomies will always be the legitimate object of close public scrutiny in one form or another.

Despite this, I strongly favor the experiment in decontrol advocated by the President, although if deregulation of rates is to be undertaken, it should go all the way. Otherwise, the distortions would be worse than at present and existing administrative machinery even more burdensome on the remaining traffic subject to regulation. But whatever the extent of decontrol, I have few illusions that the results will approximate those postulated by the competitive or workably competitive models. If our hunches regarding the level and shape of the cost functions of the various modes have much validity and if the ubiquity of *de facto* competition in transportation is less than recent assertions would have us believe, the results will be far removed from any reasonable vision of optimality. In short, I am not convinced that the effects of abandonment of rate regulation would result in any significant reduction in economic waste but welcome the chance to be proved wrong. As Ricardo might have put it, "the opinion entertained by those who believe that deregulation may not resolve the transportation problem is not founded on prejudice and error, but is conformable to the correct principles of political economy."

## DIRECT REGULATION AND MARKET PERFORMANCE IN THE AMERICAN ECONOMY

By RICHARD E. CAVES  
*Harvard University*

Perhaps one-fifth of United States national income originates in industries subject to some direct regulation, and yet economists know very little about how regulation affects the market performance of an industry. The preambles of regulatory statutes hardly provide a reliable guide. Neither does the intensity of the complaints of regulated businessmen. Economic theory can offer essential hypotheses about the nature and direction of the impact of regulation, but hardly permits us to draw detailed, quantitative conclusions on an a priori basis. Factual and statistical evidence abounds, but we have failed to give it a thorough and sophisticated review. I shall argue that the right questions have not been asked about the effects of regulation, nor the right tests performed. This paper seeks to point out the possible roads leading toward fruitful conclusions and to mention the tales told by a few travelers along the way.

### *Market Performance and Direct Regulation*

Economists studying unregulated industries in the economy now seem to have reached a reasonable degree of agreement on the major dimensions of an industry's performance and have made some progress at measuring them. The amount of resources in an industry should be just large enough that the marginal value of its output equals its long-run marginal cost. This basic condition for proper resource allocation we normally test by examining rates of return. Resources within a sector of the economy should be combined efficiently, with plants large enough to exhaust available economies of scale, technology and input combinations chosen optimally, and horizontal and vertical integration pushed far enough to exhaust economies available in these directions. Enough resources (and no more) should be devoted to sales promotion and providing information to consumers. Enough resources (and no less) should be devoted to the pursuit of innovations in products and technologies.<sup>1</sup> Finally, some would add performance tests relating to the adaptability of an industry's market prices and investment rate in regard to national goals of stabilizing employment, promoting growth, and avoiding inflation.

Not only are these performance dimensions now fairly well identi-

<sup>1</sup> For a complete account, see [2], Chaps. 9, 10.

fied; we have also gone some distance at measuring their achievement in particular cases, although the going level of sophistication in detecting efficient resource allocation obviously lies above that for measuring achievements in other aspects of performance. Serious difficulties appear when we face the problem of valuing gains in these various performance dimensions at the margin. Cruel circumstances often seem to force us to choose more of one quality of performance only by sacrificing some of another. When gaining more old-fashioned efficiency requires giving up some new-fangled progressiveness, every man must consult his own indifference surface.

Now, it would seem that the questions, both normative and analytical, which economists should ask about direct regulation all revolve around the effect of direct regulation on the market performance of the regulated industry. In this country, we use direct regulation to attain better market performance than ordinary market forces would produce. Direct regulation of an industry secures, by accident or design, a quality of market performance different from what some alternative means of social control would produce. The appraisal of direct regulation of an industry depends on measuring the difference between its going market performance under regulation and its potential performance under some different regulatory standard or alternative system of social control.

Perhaps the assertion that direct regulation aims to improve market performance should not be passed off without some discussion. As a proposition of welfare economics, this view has a clear-cut logical basis. In the American economy, we treat the attainment of "workable competition"—whatever we mean by that slippery term—under the anti-trust laws as our primary means of securing satisfactory performance in individual markets and thereby acceptable efficiency, stability, and progressiveness in the economy generally. But not all markets can provide acceptable performance by this route, and in cases of "market failure" we substitute a variety of alternative means of control. In some instances, direct regulation comprises a comprehensive set of controls over prices, price structures, the entry and exit of firms, and the quality and distribution of output, which we associate with traditional public utilities regulation. In other cases, we displace market forces less fully by employing public intervention only to limit entry, to control the quality of the service rendered, and the like. Determining where market failure has occurred and where market forces should be supplemented or replaced by some form of public control depends critically upon comparing market performance in different situations, actual or hypothetical. Furthermore, the problem of deciding how completely to regulate an industry and bypass market forces depends on the same type of comparison of performance under alternative regimes of regulation.

Nor does the political history of the adoption of direct regulation in America diverge too drastically from this abstract view of the problem. Direct regulation of individual industries has been adopted or considered for many reasons, but most of these can be squared with the goal of achieving better market performance. Some regulatory goals correspond directly to the traditional standards for market performance discussed above. Preventing excessive profits and price discrimination, providing proper outputs of so-called "public goods," and promoting infant industries can all be justified by the criteria which we use for good performance in the unregulated sector of the economy. Other stated goals of direct regulation lie outside of this traditional list, seeking to attain standards of safety, norms of equity between large and small communities or large and small businesses, and the like. When the legislative process consciously installs such goals in the preambles of regulatory statutes, the reasonable course for the economist seems to be to take them as a valid social choice of performance criteria and add them to his list.

Of course, an economist may find those goals not of his own devising difficult to interpret and measure. He may argue with the weight placed by regulatory commissions and their governing statutes on various of these performance goals. Furthermore, a fringe of special-interest objectives often surrounds the working of direct regulation, later if not sooner. But in few cases do we fail to find some parallel between what public policy seeks from regulation and the function which an economist sees for it. The same sort of test for the effect of direct regulation should be applied in either case.

The problem of linking the policy decision to impose direct regulation upon an industry to the theoretical welfare goal of improving market performance raises a related analytical problem. To put it mildly, the hierarchy of economic objectives pursued by a regulatory commission is often unclear from its decisions and its authorizing legislation. At best, a list of regulatory objectives will be discernible, but not the weights placed upon them. At worst, the basic thrust of regulation toward affecting performance lies shrouded in a fog of mechanical standards and procedural details.

In such circumstances, measuring the effect of regulation on performance serves two important uses even before one reaches the ultimate question of whether or not to regulate. First, measuring the net effect of regulation on performance serves the analytical purpose of telling the economist what game the regulators are really playing. Second, it raises the possibility of rationalizing the regulatory process within its existing framework, by notifying the regulators about the extent of their success in achieving their professed goals. Studies of regulation

so far have tended to show that the net impact of direct regulation often differs considerably from what the regulators say that they are doing [6]. Furthermore, unintended and indirect consequences may have more economic significance than the desired direct ones [7, Chap. 1] [3, Part ii]. While these unintended consequences can either improve or worsen the results achieved via the direct effects of regulation, knowledge of their significance in either case holds great importance.

Note that these questions concerning the effect of direct regulation are by nature comparative. They do not just ask: "How well do regulated industries perform?" They raise the more complex issue: "What is the net effect of regulation on performance?" The word "net" refers to the differences between actual performance under regulation and performance under some alternative regime, such as no direct regulation. Only by knowing how regulation displaces the performance which would result from market forces operating within the general economic structure or environment of an industry can we measure the economic effect to attribute to regulation. Only with this knowledge can we judge the social gain or loss from relieving an industry of regulation. It is on this shoal of comparative analysis that most attempts to appraise the performance of regulation have run aground.

### *Evidence of Effects of Direct Regulation*

What does the received body of scholarship on regulated industries have to offer? We can usefully combine a catalogue of possible ways to develop a comparative appraisal of the effects of regulation with a summary of some existing evidence.

Two basic types of approach are available for the purpose: we can compare the performance of an industry under direct regulation to an ideal norm derived from the familiar optimum conditions of economic theory; or we can match the regulated industry's performance to that of another market or set of markets subject to a different regime of control. Each sort of approach has strengths and weaknesses associated with the availability of its required data inputs and the usefulness of its intellectual output. In an imaginary utopia of unlimited information and time for research, the two approaches would be used jointly in any investigation, because they do not ask or answer exactly the same set of questions.

Comparing actual performance under regulation to a theoretical optimum will work, obviously, only for those aspects of performance for which economic theory obliges us with clear-cut criteria for the optimum, and where the available statistics allow us to estimate it. Research in the cost structure of the regulated transport sector has pushed



rapidly in this direction during the past decade. The pioneering investigation of John R. Meyer and others [7] marked out the basic problem of devising the most efficient combination of transport resources to serve the prevailing demand for the movement of goods and persons. Such a model, if fully developed, shows in the divergence of actual allocations from the computed optimum the cumulative distortions due jointly to the market structure of the transport industries and to the prevailing regulatory practices.

Besides work on the transport sector, a few self-contained studies have used the technique of comparing other actual regulated sectors to a theoretical optimum. One example is William R. Hughes's appraisal of the short-run efficiency of interfirm sales of electric power in New England [5]. In this case, an independent statistical estimate of costs can be avoided because the decision rules governing interfirm transactions themselves provide suitable evidence on the degree to which interfirm sales satisfy the appropriate marginal criteria.

These blueprints for an optimal allocation of resources permit a measure of the imperfection of the performance we actually obtain. But they do not provide a direct means of distributing blame for the deficiencies between regulatory policies, on the one hand, and features of the economic structure of the regulated industry, on the other. If such comparisons with a theoretical optimum are used for more than cleaning up the details of existing regulatory practice, then supplementary facts about the industry's basic economic environment also require close analysis.

The one extensive line of appraisal of regulatory performance which we find in the existing literature uses an abbreviated version of a comparison between actual and optimal performance. I refer to the practice of drawing direct inferences from the decision rules imposed or permitted by regulatory agencies. To take a recent example, Averch and Johnson have shown that regulation taking the form of a maximum rate of return allowed on the value of a firm's total capital will induce it to tend to make uneconomic substitution of capital for other factors, if the rate of return allowed effectively restricts maximum profits and exceeds the cost of capital to the firm [1]. This hypothesis provides an economic explanation, to place beside the obvious ideological one, for the oft noted unwillingness of private electric utilities to distribute power produced in publicly-owned generating plants. Such transactions increase their level of sales without substantially raising their investment base; thus in a sense reducing their capital-output ratios and thrusting them farther away from maximum profits. Likewise, such behavior tends to support the significance of the Averch-Johnson hypothesis about the effect of regulation. Nonetheless, such analyses of

decision rules always suffer from important weaknesses, compared to investigations which make specific calculations to estimate the nature of the unattained optimum. The conclusions can indicate the direction in which performance deviates from the optimum but not the size of the resulting misallocation. The technique thus provides little help in cases where improved performance in one dimension entails sacrifices in some other dimension. Furthermore, analyses of decision rules usually make some shaky *ceteris paribus* assumptions: that the observed decision rule actually controls the behavior of the regulated firms; that a change in the rule would not also change essential data in the cases to which it is applied; and so forth. In short, they fail to make the full comparative appraisal of the impact of regulation and can detect only the direction of a possible bias away from the optimum.

The other main sort of approach to weighing the effect of direct regulation on market performance employs for comparison not a theoretical norm but rather some form of actual alternative. Any comparison between actual alternatives involves not only different methods from one which employs a hypothetical norm but also is likely to serve a different purpose in informing public policy. If we can establish somehow the structural equivalence of market situations *A* and *B*, except for differences in the pattern of regulation, then the conclusion that the performance of *B* excels that of *A* furnishes a strong presumption for dropping the regulatory regime of *A* and switching to that of *B*. By contrast, a clear-cut finding that actual situation *A* falls short of the performance implied by hypothetical optimum *A'* tells us nothing directly about the tactical maneuvers necessary to reproduce situation *A'* in the real world. In a sense, we are only pointing to the classic danger of comparing the performance of actual and utopian systems. The warning, however, seems apt at a time when strong campaigns are being mounted both to deregulate large sectors of our regulated industries and to regulate some segments of our unregulated industries having a major impact on the general price level. Both sides may be right, but only a full analysis will tell.

Appraising direct regulation by comparing actual markets on a cross-sectional basis depends fundamentally on finding satisfactory units for comparison. A likely choice is segments of the same industry located in different regions and subject to different regulatory regimes. Thus, a popular mode of comparison has been to check the rates charged or profits earned by electric utilities under state regulation against those in states lacking regulatory commissions. The well-known study by the Twentieth Century Fund [11, Chap. 4] and others [8] have suggested by simple comparisons of average revenues or tariff schedules prevailing in states with and without regulatory commissions that regulation

indeed does result in lower prices to the consumer. George Stigler and Claire Friedland have, however, called this finding into question by pointing out that a number of purely economic elements of market structure can differ as between states, and that these differences rather than the vigil of the regulatory commissions might explain the lower charges for electricity in regulating states [10]. Taking into account such major cost-determining factors as unit fuel costs, portion of generation from hydroelectric stations, and market size and density, the statistics for various years from 1912 to 1937 reveal no significant effect of regulation on average revenues from the sale of electricity. Furthermore, no statistically significant impact of regulation could be located on the amount of price discrimination against domestic customers or the rate of return earned by a small sample of firms. Nothing in the Stigler-Friedland results explains why regulation should have proved ineffective by this test, and statistical mirages are not impossible. The point remains that the proper question has been asked and a provocative answer has resulted.

Another recent cross-sectional investigation of the electric power industry which merits attention is William Shipman's inquiry into the high cost of electricity in New England [9]. This study uses as a basis for comparisons both other private utilities in the nation generally and a particular segment of the midwest where some economic characteristics of the electric power industry resemble those of New England. The results permit a calculation suggesting how much of the inferior performance of New England rests upon a lack of "integration"; i.e., the potential gains from raising managerial and production units to the size and efficiency of those found elsewhere in the country. Shipman blames about half of New England's disadvantage in the cost of electricity on these organization factors, but does not further attempt to assort the blame for inefficient organization between the regulatory process and the other economic characteristics of the New England market for electricity.

Studies of a given industry under different regimes of social control in different regions might, of course, be pressed beyond national boundaries. Foreign countries depending basically upon market economies employ a substantially wider range of social control techniques than does the United States. Planning boards and public enterprises competing with private industry immediately come to mind. A comparison of the performance of an industry under different national control systems might prove quite valuable, but I have not encountered any serious attempts. Of course, many doubts arise: Is the spirit of Australian capitalism really the same as that of the United States? Should one proceed on the hypothesis that relations between the struc-

ture of markets and their performance regularly transcend national boundaries? The only way to find out is to try.

Another sort of cross-sectional comparison for appraising the effect of direct regulation might attempt to pick industries lying outside of the regulated sector for comparison with those within. An ideal research design might call for first measuring the economic elements of the market structure of the regulated industry, adjusting as best one can for any impact of the regulatory process upon them. Then this constellation of structural elements of the regulated industry can be placed against our general body of knowledge of the structure-performance relations prevailing in the unregulated sector of the economy, to evaluate the performance which this market structure would produce in the absence of regulation. The difference from the actual performance of the industry under regulation then becomes a measure of the net impact of regulation.

Such a procedure, which I have attempted to apply to the air transport industry [3], seems sound on scientific grounds but must cross several bridges of shaky factual underpinning on the way to its objective. When a regulatory body has long controlled entry and mergers in an industry, figuring out what level of seller concentration and barriers to entry would emerge in long-run equilibrium if the regulatory agency were swept away becomes, at best, a difficult task. Then, in plugging these market-structure traits into our general inventory of information on structure-performance relations, one encounters the many uncertainties surrounding our knowledge of the latter.

Theoretical hypotheses about the structural determinants of the quality of performance in a given industry abound, but many of these hypotheses are vaguely specified and in apparent contradiction with one another. Worse yet, few of these hypotheses have been subjected to fully satisfactory and conclusive tests. Ideally, we might hope for a body of evidence which would adequately predict the quality of an industry's performance on the basis of observed dimensions of its market structure. Such a predictive machine obviously must be built up from cross-sectional (i.e., cross-industry) evidence, but we possess such evidence on only a few of the many possible structure-performance links. A large but loose assortment of additional evidence can be dredged, bit by bit, from the accumulated shelf of industry studies.

To avoid the weaknesses of our general knowledge of structure-performance relations and still use the performance of unregulated industries for comparison, one might pick as a control group some small group of unregulated industries with market structures similar to that of the regulated industry under study. If the control is not picked with great care, however, the performance comparison becomes meaningless.

If one were to conclude that the electric power industry performs better than the cotton gray goods industry, that would hardly afford a *prima facie* case for extending direct regulation to the latter.<sup>2</sup> But picking industries with closely comparable market structures, if it is done right, is probably as much trouble as making a comparison with the general run of structure-performance relations. And, needless to say, every industry is to some degree unique, so that no twin may turn up even after the most diligent search.

### *Institutions and the Quest for Knowledge*

The mode of research proposed here into the net effects of direct regulation promises few easy conclusions. It can hardly be ruled a dead end, however, without more exploration than it has had. One indeed wonders why the travelers on the road have been so few.

Without seeking to pin blame, we might reasonably ask why so little effort seemingly has been aimed at appraising the net impact of regulation. The institutionalist heritage in the field of public utilities may explain part of the situation. In reaction to the abstract classicism (neo- and otherwise) passed off as descriptive and normative economics in late nineteenth-century America, the institutionalists rightly emphasized the structural peculiarities of actual markets and the legal and technological environment which lay behind them. The field of public utilities proved a natural subject for this approach, first as a field for reform in the campaign for regulatory commissions, and then as a field for analysis of the interaction of legal, technological, and market forces. But the institutionalist insights into the range of factors bearing upon the performance of regulated industries were by and large never carried beyond assertions of relevance to the level of measurements of influence. The many anomalies of the regulatory process itself provided ample grist for the mill of scholarship; reforming the regulatory process has often seemed so urgent a task that the question of how efficient direct regulation would prove, once whipped into fighting trim, has no doubt seemed like a low-priority item for research.

Furthermore, the conspicuous clashes between regulators and regulated create a constant temptation for the onlooker to infer the impact of regulation directly from those wishes of the regulated firms which are denied by the regulatory body. But such inferences face many pitfalls—pitfalls of the same sort that endanger any effort to infer business performance from overt decision rules, such as the familiar interpretations of full-cost pricing. In the field of direct regulation, the obvious snare lies in supposing that the existence of maximum- or minimum-rate regulation guarantees profits respectively lower or higher than they

<sup>2</sup> For a well-intentioned but basically invalid comparison of this type, see [4, pp. 25-29].

would otherwise be. But the degree to which the prices that regulated businesses would like to charge could actually be translated into a different level of long-run profits depends on many factors. In addition, direct inferences of the net effects of regulation from its virtual impact tend to miss the possibility that regulation has important side-effects on performance, as well as the more dubious possibility discussed by some political scientists that regulation typically becomes subverted to the interests of the regulated.

In short, the difficulties with the existing literature on regulation lie mainly in errors of omission. The need is for an expanded frame of reference and a few new tools of analysis. Those tools, in fact, differ little from what a well-equipped practitioner elsewhere in economics might carry: a well-oiled theoretical hypothesis generator, a fine-mesh cross-sectional net, a *ceteris paribus* clause which will work in tight places, and a large measure of common sense.

## REFERENCES

1. H. Averch and L. L. Johnson, "Behavior of the Firm under Regulatory Constraint," *A.E.R.*, Dec., 1962, pp. 1052-69.
2. J. S. Bain, *Industrial Organization* (New York, 1959).
3. R. E. Caves, *Air Transport and Its Regulators: An Industry Study* (Cambridge, Mass., 1962).
4. W. S. Gramm, "The Public Utility Scissors and Economic Policy," *Land Econ.*, Feb., 1962, pp. 21-32.
5. W. R. Hughes, "Short-run Efficiency and the Organization of the Electric Power Industry," *Q.J.E.*, Nov., 1962, pp. 592-612.
6. H. J. Levin, "Regulatory Efficiency, Reform and the FCC," *Georgetown Law J.*, Fall, 1961, pp. 1-45.
7. J. R. Meyer *et al.*, *The Economics of Competition in the Transportation Industries* (Cambridge, Mass., 1959).
8. R. J. Sampson, "Comparative Performance of Electric Utilities under State and under Local Rate Regulation," *Land Econ.*, May, 1958, pp. 174-78.
9. W. D. Shipman, *An Inquiry into the High Cost of Electricity in New England* (Middletown, Conn., 1963).
10. G. J. Stigler and C. Friedland, "What Can Regulators Regulate?: The Case of Electricity" (unpublished ms., March, 1963).
11. Twentieth Century Fund, Power Committee, *Electric Power and Government Policy: A Survey of the Relations between the Government and the Electric Power Industry* (New York, 1948).

## THE EFFECTIVENESS OF ECONOMIC REGULATION: A LEGAL VIEW

By <sup>\*\*\*</sup>ROGER C. CRAMTON  
*University of Michigan*

Some time ago I came across a quotation which ever since has been weighing on my mind. It is reported to have been spoken by an aged West Coast Indian, sitting on a rock and looking out to sea, under circumstances which I do not know. It reads as follows: "Lighthouse, him no good for fog. Lighthouse, him whistle, him blow, him ring bell, him flash light, him raise hell; but fog come in just the same."<sup>1</sup>

At the risk of oversimplification, I plan to weave my remarks around three untested assertions, each of which will require some amplification. First, many regulatory actions which are of great interest and concern to lawyers have little or no economic significance. Second, under our system the exercise of regulatory power bearing directly upon private individuals will not ordinarily be tolerated unless subjected to procedural safeguards. And third, economic regulation, even where public policy is fairly clear and the regulatory task is confined, faces enduring problems which limit its potential effectiveness. In the transportation industries, where clear objectives do not exist, the governing principle is one of ambiguous protectionism, a form of economic regulation which has demonstrated characteristic infirmities.

### I

The most basic question one can ask about economic regulation is whether it makes a difference in the behavior of the regulated industry. The lawyer is tempted to answer this inquiry by condensing the voluminous mass of decisions and regulations into an orderly analysis of regulatory policies. But, as George Stigler has pointed out,<sup>2</sup> the enumeration of an endless succession of regulatory actions provides proof, not of effective regulation, but of the desire to regulate. The net effect of the busy humming of the regulatory machinery may be only to irritate entrepreneurs and to enrich their lawyers, without effecting a fundamental alteration in the state of affairs that would have existed in the absence of regulation.

A moment's thought will indicate why the economic effect of regulation is essentially independent of the content of formal regulation.

<sup>1</sup> The source of the quotation is W. L. Prosser, "Lighthouse No Good," 1 *J. of Legal Ed.*, 257 (1948).

<sup>2</sup> George J. Stigler and Claire Friedland, "What Can Regulators Regulate?: The Case of Electricity" (unpublished paper delivered to the Economics Club of the University of Michigan in March, 1963).

The regulation may prohibit conduct which no one desires to engage in or it may encourage conduct which will take place anyway. Even if the regulation deals with conduct that would take a different course in the absence of regulation, it is always possible that the objective so devoutly desired by the regulators will not be achieved. The regulatory machinery may be too cumbersome or the ingenuity of circumvention too great. In order to determine whether the observed economic behavior in a particular industry is due to the existence of regulation, the possible effect of regulation must be isolated from other factors influencing behavior.

The significance of individual regulatory actions cannot be judged by the number of lawyers or regulators engaged in the fray, the heat of the battle, or the length of the struggle. These external indicia have little relationship to the economic significance of the proceeding. It is a safe generalization that many of the most time-consuming and expensive controversies in regulatory annals have had little economic or social significance—other than as tribal rites which lend legitimacy to conduct that otherwise might be viewed as antisocial behavior. On the other hand, some regulatory actions, which may or may not be accompanied by formal proceedings, elaborate trappings, and controversial publicity, are of great economic importance.

Two separate functions of the Federal Communications Commission illustrate the differing significance of regulatory actions. The first—allocation of the radio spectrum to various uses—is a function of enormous importance that determines the manner in which this resource is used by society. The Commission decides whether or not a frequency will be made available to citizens, such as trucking companies or ham operators, who desire to utilize the airwaves to connect their far-flung activities, to while away their leisure time, or to accomplish a thousand other purposes. If a frequency is made available for a particular purpose, the nature of the frequency and the terms on which it may be used determine the cost of the desired service to the user. It is not an accident that we have three and only three television networks. The Commission, by its actions in allocating frequencies for television use, has created a situation in which there are very few metropolitan centers possessing more than three competitive television channels.<sup>3</sup> Until this situation is remedied and a fourth or fifth station added in all or most metropolitan centers, the structure of the television industry will remain unchanged.

In sharp contrast to frequency allocation, the numerous comparative licensing proceedings in radio and television conducted by the Commis-

<sup>3</sup> The relevant materials are collected and discussed in Note, "The Darkened Channels: UHF Television and the FCC," 75 *Harv. L. Rev.* 1578 (1962).



sion have little economic significance. Comparative hearings provide a bonanza to the successful applicant, dissipate the energies of regulators, and attract the attention of scholars and critics. Yet it is doubtful that they have any substantial effect on the conduct of broadcasters.

Why is this so? In part it is because the criteria for choosing the "best" applicant are not meaningful.<sup>4</sup> A second factor is that the policies applied in comparative hearings may be and are undercut by uncontested initial grants or subsequent station transfers.<sup>5</sup> But a more fundamental reason is that the economic conditions which influence the actual conduct of broadcasters remain unchanged.

Broadcasters operate their stations in order to make money. The pattern of operation which will maximize the revenues of a particular station will not vary, regardless of who is awarded the station. The licensee of a television station, for example, will obtain the most favorable network affiliation that is available to him and he will carry a large proportion of network programs. Decisions concerning which network programs will be carried and what other programs will be offered will be made on economic grounds. Experienced broadcasters faced with identical conditions will reach roughly similar results. Comparative licensing does not affect these underlying conditions; the only practical effect is a substantial delay in the initiation of additional service to the affected community.

Of course, where the number of licenses is limited some method of distributing them to qualified applicants must be established. Alternatives that would accomplish this task in a less wasteful and more desirable manner include public bidding, grant to the first applicant in point of time, and administrative discretion. Ronald Coase has argued persuasively that the market system is best adapted to determine how many frequencies should be devoted to broadcasting uses and who should be entitled to operate on each frequency.<sup>6</sup> The broadcasters and the Commission, as well as most commentators, have neither understood nor seriously considered this proposal. Their efforts have been directed at better definition of the criteria by which the "best" broadcaster is chosen. Yet, even if more objective standards can be developed, which seems doubtful, the effort expended in comparative hearings will be wasted as long as the policies applied in such hearings

<sup>4</sup>For useful critical discussions of the substantive and procedural aspects of comparative broadcast licensing, see Henry J. Friendly, "The Federal Administrative Agencies: The Need for Better Definition of Standards," 75 *Harv. L. Rev.* 863, 1055-72 (1962); Administrative Conference of the United States, Committee on Licenses and Authorizations, "Licensing of Major Broadcast Facilities by the Federal Communications Commission" (rev. draft September, 1962) (prepared by William K. Jones, Columbia Law School).

<sup>5</sup>About 90 percent of all commercial television stations authorized in mid-1961 resulted from uncontested initial grants rather than comparative proceedings. License transfer occurs on a large scale: over 10 percent of outstanding licenses are sold each year.

<sup>6</sup>R. H. Coase, "Federal Communications Commission," 2 *J. of Law and Econ.* 1 (1959).

can be easily circumvented, and, in any event, do not affect the economic conditions faced by the individual broadcaster. Any attempt to restrict these flexible avenues for prompt service and new entry is likely to be a cure worse than the disease. Individualized licensing proceedings are not the appropriate method of dealing with the fundamental conditions which influence the conduct of broadcasters. To the extent that there is agreement on policies which broadcasters should follow—such as a minimum amount of public affairs programming or limits on the degree of commercialization—such policies can be effectively and directly implemented by general regulations applicable to all broadcasters, with license revocation as the ultimate sanction for noncompliance.

A similar analysis may be applied to other areas of regulation. Some regulatory actions affect the structure of the industry or the behavior of its firms; others affect only the fortunes of the individual participants. Decisions involving the extent and nature of competition to be permitted in a particular market have considerable economic significance. But if it has already been decided that there will be three and only three air carriers (or motor carriers) in a particular market, the choice of the particular carrier, standing by itself, is unlikely to make much difference. If there are going to be only three airlines operating between New York and Atlanta, does it really matter to the public whether the third service is rendered by Braniff or Delta?

## II

As noted above, under our system the exercise of government power bearing directly on private individuals will not be tolerated unless subjected to procedural safeguards.<sup>7</sup> Legal tradition tends to be individualistic and negative in its approach to economic regulation. "Lawyers," it has been said, "focus on the fact that public officials and tribunals are going to be fallible at best and incompetent or abusive at worst."<sup>8</sup> Traditional legal principles serve to limit discretion by providing the official with as little elbow room as possible for error, whimsey, or bias. Extended hearings allow interested persons to contribute to the decider's knowledge; the written decision must elaborate the reasons for choice by reference to the legislative standard; and court review is available to insure that procedural requirements and statutory limitations are observed. One cannot speak of law in our society—or "the rule of law"—without including procedures which allow citizens to

<sup>7</sup> For the argument at this point I am indebted to the excellent piece by Louis L. Jaffe, "The Effective Limits of the Administrative Process," 67 *Harv. L. Rev.* 1105 (1954).

<sup>8</sup> Kingman Brewster, "Enforceable Competition: Unruly Reason or Reasonable Rules," *A.E.R.*, May, 1955, pp. 482, 484.

participate in a meaningful way in governmental decisions which bear directly upon them as individuals.

It is often contended, especially by nonlawyers, that judicial attitudes and processes, with their emphasis on adversary presentation of evidence, cross-examination, and contentious record-building, are unsuited to the implementation of broad public policies by regulatory agencies. There is often justification for the charge that judicialized procedures are applied thoughtlessly to functions to the solution of which they do not contribute. In my judgment, the comparative hearing in broadcast licensing is one such instance.

There is room for legitimate difference of opinion concerning the point at which the increased accuracy of formal decisions, when contrasted with summary decisions, is offset by their relative inefficiency. Yet there is little room for doubt that the citizen, regardless of relative efficiency, finds summary discretion distasteful and unacceptable when applied to his own concerns. If he is to be excluded from a particular activity or if the return which his property may earn is to be controlled on individual grounds, he demands an opportunity to participate in the making of the decision. Our legislatures reflect this democratic choice when they insist, in statute after statute after statute, that regulatory powers be exercised only after full hearings. As long as this preference prevails, the procedure which it demands will constitute a limiting condition on economic regulation.

### III

There are inherent limitations on the effectiveness of economic regulation even where public policy is fairly clear and the regulatory task, relatively speaking, is confined and manageable. The simpler case of economic regulation—the determination of maximum rates of a conventional public utility—has not been performed with obvious success. I do not assert that public utility regulation has been a failure. I do maintain, however, that unqualified assertions of its effectiveness would be unwarranted. The lesson of a half-century of experience is that the environment generates enduring problems which limit the potential effectiveness of rate regulation.

The regulation of interstate telephone rates by the Federal Communications Commission during the period 1953 through 1962 illustrates the general proposition.<sup>9</sup> The objectives are clear and limited;

<sup>9</sup> Congressional investigations have resulted in the publication of a great deal of material relating to interstate telephone regulation, including internal memoranda prepared by the FCC staff for the Commission. See *Hearings on the Antitrust Problems of the Satellite Communications System Before the Subcommittee on Antitrust and Monopoly of the Senate Committee on the Judiciary*, 87th Cong., 2d Sess., pt. 2 (1962); *Antitrust Subcommittee of the House Judiciary Committee, Report on Consent Decree Program of the Department of Justice*, 86th Cong., 1st Sess., 73-83 (1959); *Hearings on the Consent Decree Program of the Department of Justice Before the Antitrust Subcommittee of the House Judiciary Committee*, 85th Cong., 2d Sess., 3427-570 (1958).

and the methods are traditional and well established. Yet a detailed look at the methods and mechanics of regulation suggests strong doubts concerning its effectiveness: disputed issues have been compromised by the Commission after private negotiation with the Bell System; standards for determining allowable expense, items includible in investment, and cost of capital have never been determined; and a relatively high rate of return of 7.5 percent over an extended period has encouraged investors to treat A. T. & T. stock as a growth rather than as an income security. The FCC has never even explained or attempted to justify this state of affairs.

I do not offer this as a demonstration of the ineffectiveness of federal telephone regulation. My thesis is that this pattern is typical, that doubts of similar dimension could be raised concerning the performance of nearly all regulatory schemes. Why is this so? What are the limiting conditions on the effectiveness of economic regulation that emerge from the task itself or that are imposed by the environment in which it operates?

I have spoken of the relative simplicity of conventional public utility regulation in that its purposes are fairly clear and its methods well established. Public utility regulation is made a manageable function by the ability to treat each company as a separate unit subject to detailed accounting regulations. In addition, most areas of conventional public utility regulation are shielded in varying degrees from the disruptive effect of outside influences, such as competition from unregulated substitutes. Even so, the complexity of the regulatory task is staggering. The treatment of joint costs in the telephone industry or the explanation of the toll rate disparity between interstate and intrastate rates are problems that tax the abilities of able economists.<sup>10</sup> It is easy to underestimate the difficulty of the problems if one is not responsible for the results nor embarrassed by an overabundance of information.

Inadequacies of personnel and appropriations constitute a second limiting factor that seems to be endemic. A handful of poorly paid employees are asked to perform complex tasks of regulation requiring zeal and imagination. Before long nearly all of the available manpower is tied down in the processing or review of routine matters. Once the immediate needs which produced regulation have been assuaged, the public loses interest and the agency falls into a routine in which day-to-day accommodations are made with those subject to the regulation.

A third limiting factor, already mentioned, is the inertia resulting from the agencies' limited power under typical statutory provisions to take decisive action without extensive formal proceedings. Indeed,

<sup>10</sup> See the study by Leland L. Johnson, *Communications Satellites and Telephone Rates: Problems of Government Regulation* (prepared for NASA by the RAND Corp. and reprinted in *Hearings on the Antitrust Problems of the Satellite Communications System Before the Subcommittee on Antitrust and Monopoly of the Senate Committee on the Judiciary*, 87th Cong., 2d Sess., pt. 2, 603-52, 1962).

many agencies are unable to defeat proposals of others without a decision based on a formal hearing and record. Yet formal proceedings have their virtues; and their legislative popularity assures that they will be with us for a long time.

Thus far public utility and broadcast regulation have been used as vehicles for stating several rather simple theses. I now move to transportation regulation, a field where the overlay of legislative objectives and administrative authority is hydra headed and more complex. This is the arena of "ambiguous protectionism": that unnatural state in which the salt of competition is joined in uneasy mixture with the pepper of protectionism and the whole porridge seasoned with the paprika of subsidy and promotion. Experience with various forms of ambiguous protectionism over several decades is now sufficient so that some of its characteristic infirmities can be stated: (1) lack of a clear legislative mandate tends to stultify administration; (2) the passive functions of protectionism tend to be implemented more effectively than the affirmative functions of development and coordination; and (3) competition from alternative forms of transport that are exempt, unregulated, or promoted by another agency tends to limit even the effective performance of protective functions.

1. The objectives of ICC regulation of domestic surface transportation are typical in that they reflect several disparate objectives. The earliest objectives—the protection of shippers from unreasonable charges and unjust discrimination—have a procompetitive flavor, but they have become less important as competition has tended to replace monopoly in transportation. More recent regulation is based on the premise that unrestrained competition in the transport field will produce harmful results; transport companies must be protected from their own mistakes and their competitors' greed. The Commission was given a vague mandate to displace carrier judgments in order to protect carrier revenues, coordinate the transportation system, and preserve certain historic policies which favor the broad movement of commodities, affect the location of industry, and allow or require the continuance of unremunerative services.

After the first years of regulation, when the initial enthusiasm has been replaced by a convenient reliance on routine solutions, the vague mandate—expressed in terms of some undefined "public interest"—is likely to produce a timid and unimaginative approach. Agency members, who are readily identifiable and exposed to attack, are reluctant to assume tasks of national planning which they or others may feel are beyond their competence or commission. Lack of a clear legislative mandate stultifies administration.

2. The so-called "independence" of the regulatory agency is a source

of weakness when the agency is not implementing policies that find strong support in a democratic consensus. Isolated from the sources of political power, forced to evolve working arrangements with those it regulates, dependent on Congress for funds and on the President for reappointment, and harassed by an unending succession of congressional investigations and industry pressures, the agency withdraws from committing itself on decisive issues of policy. It drifts along, responding to the most urgent pressures as they arise and perpetuating, for the most part, regulatory patterns which were created in the past to meet different problems. In short, the agency becomes passive, backward-looking, and resistant to change.

An important implication of these general tendencies is that the more passive functions of protectionism are more effectively implemented than the affirmative functions of planning, development, and coordination. The ICC, for example, has been notably ineffective in achieving affirmative goals of coordination and efficiency. Its powers and procedures and the practicalities of their use limit it to a negative and restraining role. If it desires to restructure the transportation system to ameliorate the problems of the New Haven, it is cast in the role of passing on a merger of the New York Central and the Pennsy—not exactly the proposal that it was likely to have had in mind. The ICC, by expending most of its resources in formal hearings devoted to such matters, can negative carrier proposals which threaten the *status quo*; but, like other transportation agencies, it lacks the powers, resources, or inclination to formulate and implement affirmative programs. The Commission, understandably, concentrates its energies on the most manageable and specific of the tasks assigned to it by the legislature: protecting interests created in the past and moderating the effects of undesirable change. The Commission's attempts to prevent or ameliorate departures from the existing rate structure and traffic pattern—a kind of soft-hearted and backward-looking cartelism—indicate its true purpose and function as a conservative body fighting a rearguard action against the inevitable forces of change.

3. It is ironic that even the regulatory function that is performed most effectively in the transportation field—that of protecting established carriers from new competition—is hedged about with statutory provisions and jurisdictional limitations which impair its effectiveness. The provision for “grandfather rights” to carriers in business when the Motor Carrier Act was passed meant that the existing structure of the industry, which allegedly was characterized by instability and oversupply, would be perpetuated.<sup>11</sup> Moreover, each carrier possesses an

<sup>11</sup> See Note, “Federal Regulation of Trucking: The Emerging Critique,” 63 *Colum. L. Rev.*, 460 (1963).

unlimited right to expand by adding new equipment and service along its certified routes. With a large number of firms already in existence, control of the number of trucks would be necessary in order to remedy the purported evil of chronic oversupply. Nor is the Commission granted any power to rectify mistakes made in granting operating rights; licenses can be revoked only for repeated and willful violations.

Absence of regulatory controls over local carriers, private carriers, and haulers of agricultural commodities—together constituting more than half of truck transportation—means that at best ICC policies can be only partially effective. Finally, the relatively free transferability of operating rights allows new blood to enter the industry and provides an opportunity for expansion on the part of aggressive operators.

These limiting factors suggest an argument that transportation regulation may be tolerable because of its ineffectiveness. There has always been a substantial amount of competition within each transportation industry. The division of responsibility among a number of jealous agencies, each adopting a paternal attitude toward the carriers under its wing, assures strong intermodal rivalry. And in any event, the availability of unregulated alternatives—exempt, unregulated, and private carriage—limits the market power of the regulated common carriers. These circumstances introduce a substantial dose of free market discipline into transportation markets. The record of performance in the transportation industries justifies the assertion that the present half-way house, despite its deficiencies, is superior to the more effective protectionism that would be the only likely result of centralized control and broader authority over the transportation industries, the proposal that is concealed under slogans of the need for increased “coordination” and “integration” in transportation.

But surely it is a counsel of despair to argue that this great nation has blundered its way into a compromise between competition and regulation which is tolerable because other alternatives might be worse. The existing regulatory schemes undoubtedly have produced a substantial misallocation of resources. Public aids have distorted the terms of intermodal competition. Regulatory support of high common carrier rates has contributed to the dramatic growth of private and unregulated carriage. Attempts to preserve the existing pattern of traffic allocation have prevented the railroads from full utilization of the economies of scale of their existing capacity. Trends toward concentration through merger and toward semicartelization through the activities of rate bureaus have been encouraged.

Economic regulation, like other tools of governmental control of society, has characteristic virtues and vices. When the objectives to be achieved are clear and precise, techniques for their realization can

usually be designed. The task is not easy and the results are not fore-ordained. But the job can be done. When the blunderbuss, however, is substituted for the rifle and a potpourri of vague or inconsistent objectives is thrown into regulatory safe-keeping, the results are likely to be limited to the protection of established interests or to be wholly unpredictable. Regardless of defects or unintended consequences, a regulatory scheme is likely to generate interests that make deregulation extremely difficult. Increased attention to the problem of accomplishing discrete regulatory objectives in the most limited fashion, disturbing other relations as little as possible and avoiding unintended consequences, would pay great dividends. The lighthouse can and does perform essential tasks; but it cannot be expected to keep out the fog.



## DISCUSSION

ERNEST W. WILLIAMS, JR.: The three papers all treat of regulation, but from different points of view. Wilson seeks to appraise the effect of regulation upon resource allocation in transportation. Cramton seeks to test the efficiency of regulation in certain subject-matter areas. Caves poses questions with respect to the market performance of the regulated industries and seeks to explore methods by which answers may be sought.

Wilson sets forth succinctly the principal criteria employed by the Interstate Commerce Commission in rate regulation. I see no reason to quarrel with the general picture he creates. He is right that regulation, even if only by virtue of its procedural requirements, tends to deter predatory pricing. He has made a contribution in redirecting attention to the difference between marginal cost of sales and of output units in transportation and its significance—a point which he has elaborated elsewhere. And, in the light of the varied service characteristics exhibited in practice, he is correct in distrusting the adequacy of the central tendency results which are the product of empirical cost analysis.

In his discussion of regulatory attitudes toward discrimination he ignores, although he has elsewhere recognized, the tendency for the relatively uniform treatment of distance despite varying conditions of density, terrain, and other factors which affect cost to create a "reverse" discrimination akin to the type imposed by reluctance to countenance reduced rates for shipments in excess of carload lots. It is possible departures of rates from cost are of even greater significance in this type of situation. And it is clear that territorial or system average costs can serve only as a point of departure in the appraisal of specific hauls. Although this aspect of the pricing problem has received much emphasis abroad, particularly in France, it is little discussed in our literature.

It is, perhaps, surprising that in his discussion of the share-the-traffic criterion he does not call attention to the consequences of enforcing a division of traffic between competing modes when the differences in cost levels are sharp. Under these circumstances, though aggregate net realized by all participants in the traffic may be held at a level approaching the practicable upward limit, the cost of performing the service has also been held to a level higher than would result from a concentration of the traffic in the hands of the low-cost mode. In practice, regulatory attention seems most often to be focused upon the prospective behavior of gross revenue, rarely upon net.

It seems plain that there is considerable misallocation within the transport system. As Wilson suggests, however, we are in no shape to measure it nor to assign a share of blame to regulation. In consequence, the case for deregulation is to be made on grounds of the political and administrative impossibility of complete and effective regulation and the artificial restraint which incomplete regulation imposes on regulated carriers with the result that relative overexpansion of unregulated carriage is induced. Yet we must approach deregulation without confidence that the results would be an improvement on

the present position. While the effects of deregulation remain so largely speculative it is doubtful that major revision of the law can be secured.

Caves concentrates upon this very problem and suggests that students of regulation have seldom asked the right questions, rarely applied the right tests. He suggests that the appraisal of direct regulation depends upon measuring the difference between the regulated industry's market performance under regulation and its potential performance under some other regime. A considerable part of his paper is devoted to a discussion of methods by which such measurement might be attempted. It is eloquent upon the difficulties which assail the researcher who seeks to shed light on the question. Doubtless those difficulties account in no small measure for the fewness of the attempts which have found their way into the literature. To say, however, that those most deeply concerned with the regulated industries have not asked questions as to the comparative behavior of those industries under various schemes of regulation as well as in the absence of direct regulation is to ignore the genuine concern which many students, especially of transportation, have felt about the regulatory issues over at least the past three decades.

In view of the fact that a substantial portion of transportation is conducted outside the impact of direct regulation, it might appear that a logical beginning would be a study of comparative behavior of regulated and unregulated segments of the motor and water carrier industries. Although there are differences in scope of operations which might well necessitate qualification of the results of such study, the degree of confidence with which hypotheses could be offered should be greater than where diverse forms of transport are studied or comparisons made with nontransport industries. Unfortunately, the unregulated sector leaves no trail of recorded information available for study, so that the researcher is faced as a first step, with the problem of identifying the firms which may be taken to represent it even in advance of laying on a program to collect his data. In transportation, moreover, the largest part of regulatory activity and the part which may be supposed to have the greatest impact upon market performance concerns the competition between the several forms of surface freight transport. More uncertainty is expressed concerning the probable behavior of the railroad industry under modified regulation than any other. Here there are no unregulated segments open to examination and the period when the industry was substantially unregulated is sixty years back in our history.

Caves is certainly right that the normative approach sheds little light upon the question what performance under a different regulatory scheme would be. It is also open to the many difficulties which Wilson notes as standing in the way of confident measurement of marginal costs on the one hand, and of assessing the impact of nonregulatory factors on the other. Yet it can afford at least a rough standard for appraisal of both the private and public forces at work and supply a framework within which analysis rather than mere description of the regulatory performance may conceivably be undertaken. Here Meyer and associates have made a contribution, but it remains for someone to undertake an approximate estimate of the level of transport cost which the economy might face given: (*a*) reasonably optimum allocation of the traffic among forms and (*b*) effective organization and performance within each of

the forms. Such an estimate would give us a rough idea of the magnitudes in issue.

Time does not permit discussion of Cramton's paper which, however, raises substantial questions about the significance as well as the efficiency of the regulatory process. The three papers taken together disclose a wide ranging disposition to question the effectiveness of regulation in improving economic performance within the regulated industries.

RONALD H. COASE: The views expressed in the papers presented in this session seem, where they overlap, to be broadly in agreement. What the regulatory commissions are trying to do is difficult to discover; what effect these commissions actually have is, to a large extent, unknown; when it can be discovered, it is often absurd. This grim report gave me a good deal of satisfaction, and for the only reason which really weighs with a scholar: it would appear to be true.

When we discuss the effects of regulation we are, of course, interested in much more than in what ways the subsequent course of events would have been different if this particular commission had not been set up or that statute passed, and the pre-existing conditions had been allowed to continue. We are also interested in evaluating these differences in order to come to a conclusion as to whether the original decision was wise. But if we are interested in economic policy, it would clearly be a mistake to confine ourselves to this question alone. We should also be interested in other possible modifications in the system, with, for example, what would happen if the powers of a commission were enlarged or narrowed or changed since this might well produce a state of affairs better than either now exists or would have existed if the commission with its present powers had not been established. It is in this broad sense of an enquiry into the effects of a whole range of regulatory arrangements that I conceive our task when we talk about discovering the effects of regulation.

Professor Caves's main concern is with the method to be used in ascertaining these effects. With the direction in which Professor Caves's thoughts are moving, I am extremely sympathetic. My objection is that he does not go far enough. It is perhaps natural that Professor Caves, coming from Harvard, should be too conservative and that, although perceiving the truth, he should show himself reluctant to abandon existing habits of thought. Professor Caves distinguishes two possible approaches. With the first we compare the performance of the regulated industry with "an ideal norm derived from the familiar optimum conditions of economic theory." In taking as his first approach a comparison with an optimal situation, Professor Caves is selecting an approach commonly used by economists when dealing with questions of economic policy. It is, in most cases, a thoroughly bad approach. It is obvious that if you are comparing the performance of an industry under regulation with what it would be without regulation, there is no reason to assume (indeed there is good reason not to assume) that either of these situations will correspond to anything an economist would call optimal. The same is true if one is thinking of modifications in the system. None is likely to be optimal

since it is quite certain that, whatever may be the characteristics of the ideal world, we have not yet discovered how to get to it from where we are. Contemplation of an optimal system may suggest ways of improving the system, it may provide techniques of analysis that would otherwise have been missed, and, in certain special cases, it may go far to providing a solution. But in general its influence has been pernicious. It has directed economists' attention away from the main question, which is how alternative arrangements will actually work in practice. It has led economists to derive conclusions for economic policy from a study of an abstract model of a market situation. It is no accident that in the literature (and for that matter in Professor Caves's paper) we find a category "market failure" but no category "government failure." Until we realize that we are choosing between social arrangements which are all more or less failures, we are not likely to make much headway. And in other ways this approach is harmful. "The amount of resources," Professor Caves tells us, "should be large enough that the marginal value of its output equals its marginal cost." This is, of course, blackboard economics, in which with full knowledge of the curves (which no participant in the actual economic process possesses), we move factors around (on the blackboard) so as to produce an optimal situation. This may well be a good way of teaching the tools of economic analysis but it gives students a very poor idea of what is normally involved in deciding on economic policy. The kind of question which usually has to be decided is, for example, whether the administrative structure of an agency should be changed or a certain provision in a statute amended. That is to say, what we are normally concerned with are social arrangements and what is economically relevant is how the allocation and use of factors of production will change with a change in social arrangements. There is little that we can learn about this from a study of theoretical optimal systems.

Given these views, it will not surprise you to learn that I found Professor Caves's second approach—the comparison of a regulated industry with an industry not subject to regulation—more congenial. It does focus attention on the actual working of alternative arrangements. The only drawback to reliance on this approach is that we will very rarely have an opportunity to use it. As Professor Caves mentions, it is hard to find industries which are otherwise comparable apart from the kind of regulation.

Is there an alternative? I think there is, although what I have to say here cannot go beyond assertion. I believe that by a detailed study of an industry or organization it is possible to obtain sufficient understanding of how it operates to be able to say how its performance would be affected by changes in circumstances; for example, the introduction of a particular form of regulation. I think that to study other industries or organizations which have been subject to similar regulation would enlarge our understanding and would assist in forming such a judgment. But a mechanical comparison of A (with the regulation) and B (without) is not likely to be helpful since the effect of the regulation in A need not be the same as it would produce in B. What is needed is an act of imaginative reconstruction. This must, however, be based on detailed knowledge and such knowledge can only come slowly as

the result of the work of many scholars over many years. But this should not cause us to despair. Ours is not an enterprise in which everything has to be sacrificed for speed. After all, we are not trying to get to the moon.

Professor Caves raises the question of why this branch of our subject is in such poor shape and suggests as a partial answer the baleful influence of the American institutionalists. This particular question is of more interest to Professor Caves than to me, since he thinks there are branches of economics which are in good shape. But I would like to make a short comment on his answer. Their influence cannot have been bad because they were institutionalists. It can hardly have been bad because they were American. Perhaps Professor Caves had in mind that they were not theorists. If so, I would like to say that we have less to fear from institutionalists who are not theorists than from theorists who are not institutionalists.

The legal framework is perhaps the most important institutional factor which economists interested in securing the kind of regulatory agency and the kind of regulation which would promote economic efficiency should take into account. It was therefore a most happy thought on the part of the organizers of this session to have invited a lawyer to address us and we were most fortunate in the choice of Professor Cramton. His paper will, I am sure, have a most salutary effect on our thinking about the regulatory process. Professor Cramton quoted a statement in which it was said that lawyers "focus on the fact that public officials and tribunals are going to be fallible at best and incompetent or abusive at worst." This is not the way in which economists habitually think about governmental bodies. We tend to think of them as benevolent associations waiting to take over when the Invisible Hand points in the wrong direction. As I remarked earlier, in economic analysis we have "market failure" but no "government failure." A lawyer's experience may perhaps lead him to take too jaundiced a view of human nature; but his experience affords a welcome corrective to what is implicitly assumed in much economic discussion. I thought Professor Cramton's discussion of procedural matters was interesting in much the same way. He explained why it is, with our existing legal and political framework, that regulatory commissions cannot be expected to act with great speed and why adjustment of their actions to changed economic conditions or knowledge is bound to be difficult. This again is a factor to be taken into account by an economist in discussing the economic role of the regulatory commission.

I now turn to the more purely economic questions considered by Professor Cramton. He argues that while some of the actions of regulatory commissions have important effects on the allocation of resources, others, to which they devote a good deal of time and energy, do not. This he illustrates from the work of the Federal Communications Commission. His account is quite correct. But this should not lead us to conclude that the use of a procedure which involves comparative hearings is without economic significance, and for a reason which Professor Cramton recognizes. It is true that the allocation of resources within the broadcasting industry is little affected. But economists are interested more generally in the use of resources and there can be little doubt that the present system involves a waste of resources. The main

resource that is wasted is the time and energy of lawyers. But Professor Cramton has surely convinced us (though this was obviously not his intention) that the talent of lawyers is a scarce resource of great value. That the present system wastes this resource is a significant economic effect.

The purpose of Professor Cramton's discussion of the effectiveness of regulation is to show that, even if we confine ourselves to the simpler tasks, or what are conceived as such (the determination of the general level of prices or the rate of return), regulation has not been an obvious success. I would not regard some of the instances given by Professor Cramton as decisive. But what is important is that Professor Cramton's examination has left him uneasy and that we are not in a position to say how far his doubts are justified. This is something which economists should take very seriously.

I have one final point. Professor Cramton ascribes some of the ineffectiveness of the commissions to the difficulty of the economic problems which they face. This is no doubt true. But these problems are difficult because economists have not yet learned how to solve them. Part of the poor performance of the regulatory commissions (and Professor Caves made essentially the same point) must be attributed to "economists' failure." I trust that we will all be as anxious to remove this failure, which it is perhaps within our power to do, as those other failures, which we should recognize in our work but which, in general, it is not within our power to remedy.

# EFFICIENCY IN THE USE OF NATURAL RESOURCES

## ATTAINMENT OF EFFICIENCY IN SATISFYING DEMANDS FOR WATER RESOURCES

By IRVING K. FOX and ORRIS C. HERFINDAHL  
*Resources for the Future, Inc.*

What constitutes an efficient allocation of productive services depends on the objectives of the system. Our discussion is confined to the problem of how an efficient allocation of water resource investment can be approached more closely, given that the objective is to increase national income.<sup>1</sup> The subject of this paper is a large one and numerous studies bear upon it,<sup>2</sup> but our treatment must, of necessity, be limited in scope. Therefore, we have restricted ourselves to an assessment of the results of existing practices within the federal government upon efficiency. This we seek to do by: (1) Examining certain key federal policies which bear on efficiency. (2) Analyzing certain features of 178 reports on projects authorized in 1962 by the Congress for construction by the Corps of Engineers.<sup>3</sup> These projects involve an initial capital investment of about 3 billion dollars. (3) Comparing in a limited way the Corps projects authorized in 1962 with Corps projects authorized in 1950.<sup>4</sup>

Federal appropriations for water resources and related activities were \$1.5 billion in fiscal year 1963. The Corps of Engineers, which now receives approximately two-thirds of the total federal appropriations for water resources, estimates that its expenditures will increase about 6 percent a year from approximately \$1.0 billion dollars in 1963 to about \$2.5 billion in 1980.<sup>5</sup> Of greater significance than the size of the investment is the fact that federal programs dominate the pattern of development on major streams and their tributaries. The structures erected under these programs alter for the indefinite future the hydrology of

<sup>1</sup> Other objectives have been suggested or appear implicit in some public water programs. See Arthur Maass, Maynard M. Hufschmidt, *et al.*, *Design of Water-Resource Systems* (Harvard Univ. Press, 1962), Chap. 2. However, to date public agency economic evaluations have been concerned primarily with the increase in national income.

<sup>2</sup> Among the books concerned with this subject are: Otto Eckstein, *Water Resource Development* (Harvard Univ. Press, 1961); John V. Krutilla and Otto Eckstein, *Multiple Purpose River Development* (Johns Hopkins Press for Resources for the Future, 1958); Jack Hirschleifer, James C. DeHaven and Jerome W. Milliman, *Water Supply: Economics, Technology, and Policy* (Univ. of Chicago Press, 1960); Maass, Hufschmidt, *et al.*, *ibid.*; and S. V. Ciriacy-Wantrup, *Resource Conservation: Economics and Policies* (Univ. of California Press, 1963).

<sup>3</sup> Public Law 87-874.

<sup>4</sup> Public Law 81-516.

<sup>5</sup> Speech by Lt. Gen. W. K. Wilson, Jr., Chief of Engineers, U. S. Army, "Water Resources Development and National Defense," delivered May 18, 1962, National Rivers and Harbors Congress.

the nation's river systems and thus determine how their waters and related lands may be used for a long time to come.

The attention which has been given to the economics of water use has resulted in the application of reasonably sophisticated techniques of economic analysis in the design of water resources projects, and there can be little doubt that the standards and procedures now in use are effective in preventing the construction of many uneconomic projects. The advances which have been made since World War II are substantial, and federal employees responsible therefor merit commendation for this accomplishment. But progress in application has been quite uneven. As this study reveals, there remains substantial room for improvement.

### *I. The System for Investment Decision Making*

The structure of interest group relationships which generates decisions on water investments by federal agencies is an elaborate one. Certain features of this system are of key importance in assessing economic efficiency.<sup>6</sup> Perhaps the most important is that the beneficiaries of a federal water investment project commonly bear only a small portion of the costs of the project. For projects planned and constructed by the federal government, the major part of the cost will be met from federal taxes, although most of the benefits will accrue to the region. The Corps projects authorized in 1962 involved a total initial investment of \$3 billion; 90 percent of this was designated as federal investment. On the basis of a crude estimate that assumes full repayment of power and municipal and industrial water costs, the ultimate federal cost would be about two-thirds of the total cost. The quite natural result is that the agency allies itself with local and regional organizations representing those who will benefit; e.g., contractors, landowners, shippers, construction materials firms, and local merchants. The interests of both the agency and the local beneficiaries will be served by an expanding program. On the one hand, this interest fosters optimistic estimates of benefits and costs.<sup>7</sup> On the other hand, it generates a powerful

<sup>6</sup> For a more detailed description of these procedures and their legal foundation, see Chap. 10, *Report of the President's Water Resources Policy Commission*, Vol. 3, *Water Law* (GPO, 1950).

<sup>7</sup> For studies of benefit estimation see Fred A. Clarenbach, "Reliability of Estimates of Agricultural Damages from Floods," *Task Force Report on Water Resources and Power*, Vol. III (Committee on Organization of the Executive Branch of the Government, GPO, 1955), and Edward F. Renshaw, *Toward Responsible Government* (Chicago: Idya Press, 1957), esp. pp. 72-78. For studies of cost estimation see *The Role of Uncertainties in the Economic Evaluation of Water Resources Projects* by Edward G. Altouney (Institute in Engineering-Economic Systems, Stanford Univ., 1963) and House Committee Print No. 21, 82d Cong. Underestimates of cost are of two kinds. One is the underestimate of the eventual scope of the project at the time of initial authorization. Print 21 indicates that this accounted for a 60 percent increase in costs of Bureau of Reclamation projects and a 35 percent increase in costs of Corps projects. The data on which the Altouney study



joint effort to secure favorable congressional action on the projects decided upon. Other state and federal agencies may have a moderating influence by securing greater consideration for the specific interests they represent, but they have no effect on either the design or the estimates for those parts of the projects that are outside these interests.

When the project gets to the Bureau of the Budget in Washington, that Bureau is not in a position to alter the decision except in a relatively minor way if, in fact, there has been success in generating strong political support for the project. Indeed, the President may welcome the opportunity to give his support to the project in return for political support of some other measure of vital interest to him. In this way, a very dubious project may turn out to be "consistent with the President's program," as the phrase goes. Congress is not in a position to take adverse action if political support for the project is significant. The congressmen and senators representing the beneficiaries are in an awkward position to oppose the project, and other congressmen and senators cannot easily take adverse action without running the risk of reprisal when projects in which they have a direct interest come up for consideration.

It is abundantly clear that those who benefit from federal water resources projects dominate the investment decision. The primary restraint is the professional integrity of the agency personnel who plan the project. Even here, the decision-making environment is such that it would be surprising if agency personnel could be completely objective. It is quite possible that the major departure from efficiency in federal water resources programs stems from optimistic estimates of benefits and the neglect of cheaper alternatives by conscientious federal employees who unfortunately function in a system in which most of the pressures are in one direction. We emphasize that these deficiencies are in no way a reflection on the integrity or competence of federal employees but instead are attributable to the policy framework within which they function.

It is important to recognize that minor adjustments in estimates at some stages of the evaluation can result in large changes in estimated benefits because of leverage effects. For example, if rent on agricultural land is 40 percent of gross receipts, a 1 percent increase in the receipts estimate means a  $2\frac{1}{2}$  percent increase in rent, viewed as a residual.<sup>a</sup>

---

of Bureau of Reclamation projects is based indicates that this factor accounts for a 94 percent increase. The other underestimation is attributable to errors in estimates of cost given the scope of the project. Print 21 indicates that this accounted for a 10 percent increase in costs of Bureau of Reclamation projects (excluding the Missouri Basin) and an 18 percent increase in Corps costs. The Altooney study indicates a comparable figure of 18 percent.

<sup>a</sup> For a more thorough discussion of the political and administrative process for water investment decision making, see Maass, *et al.*, *op. cit.*, Chap. 15.

## II. *Benefit-Cost Evaluation Practices*

The scope of this paper permits only a brief discussion of a few of the key issues involved in an appraisal of federal practices for evaluation of benefits and costs of water resources programs.

*Application of the Marginality Principle.* It is the expressed policy of the federal water resources agencies to undertake only those activities and increments thereof for which the incremental benefits exceed the incremental costs.<sup>9</sup> That is, the benefit-cost ratio, assuming proper measurement of alternative costs, must be over one for the whole project and for each of its subprojects. There will be a net loss if substandard projects are combined with justifiable projects. If project No. 1, yielding positive net benefits of 10, is combined with project No. 2, yielding a net benefit of  $-5$ , the total is  $+5$  all right, but the net gain to the economy can be increased to  $+10$  simply by transferring productive services from project No. 2 to other uses where the value of their marginal product will equal their costs.

A proper application of the marginal principle requires something more, however: marginal net benefits must be zero. This means that it must not be possible to increase total net benefits by making the size of the project larger or smaller. For example, the dam must be neither too high nor too low; the channel must not be too deep or too shallow.

We are not aware of any recent case where the costs of a system of projects exceeded benefits as calculated by the planning agency. It is true, however, that every one of the participating projects (not including the storage units) of the Upper Colorado River Storage Project (total cost about \$1 billion) planned by the Bureau of Reclamation has an incremental benefit-cost ratio of less than unity based upon direct benefits.<sup>10</sup> And our examination of the 1962 project reports of the Corps of Engineers revealed one case where it was clear that marginal net benefits were negative.<sup>11</sup>

Ordinarily it is impossible to tell whether the size of a project is correct, however, since data on the costs and benefits of marginal increments usually are not given in project reports. The one exception evident in the project reports of the Corps of Engineers is for incremental changes in navigation channel depth. But in general, engineers seem to be motivated to build projects which utilize the "full physical potentialities of the site" (i.e., build to the point where incremental costs rise sharply) rather than to be governed by a comparison of incremental costs and benefits. Earlier studies have reported that the Corps of Engineers followed this practice in the design of flood control projects, thus

<sup>9</sup> S. D. 97, 87th Cong., 2d Sess.

<sup>10</sup> See table facing p. 21 of Sen. Doc. 101, 85th Cong.

<sup>11</sup> Sen. Doc. 127, 87th Cong., 2d Sess.

tending to carry protection beyond the point where net benefits are maximized.<sup>12</sup> An examination of current practice which would compare marginal costs and returns would be very helpful.

*Selection of Discount Rates.* It is not within the scope of this paper to examine the conceptual aspects of choosing a discount rate for federal water projects, but it should be noted that there is widespread agreement among economists that from an efficiency point of view a rate is required which is more comparable with private rates than those used currently or in the past.

For the projects authorized in 1962, the rate generally applied to federal investment costs was 2½ percent, but a higher rate often was applied to that share of the initial investment borne by nonfederal interests.<sup>13</sup> Also, in calculating the annual equivalent of the rising portion of a benefit stream, a higher rate of discount has often been applied.

If a discount rate of 4, 6, or 8 percent had been applied to those investment costs for which a lower rate was used in the project report, the following percentages of initial gross investment would have a B-C ratio of less than unity: 9 percent, 64 percent, and 80 percent, respectively. Since our calculation assumed the annual equivalent of benefits in the project report to represent the true benefit stream, the above percentages understate the effect of a higher rate of discount where the project report rate of discount used on the rising benefit stream was lower. Our calculations assume that the data and estimates in the project reports are correct, of course. Needless to say, application of the higher rates would not have eliminated all of the projects having a new B-C ratio of less than unity but would have reduced the scale of development in many instances or stimulated attention to other and cheaper ways of attaining the same objectives.

Involved in the selection of an appropriate discount rate is the question of uncertainty stemming from technological change and major unexpected modifications in the character of demand. Experience with both of these factors underscores the desirability of maintaining flexibility in meeting future demands. This can be encouraged in a rough and ready way by such means as a higher discount rate or a shorter assumed life for projects. It is also important that serious consideration be given to project designs which retain flexibility of action, for example, by using general purpose rather than specialized design or by using less durable structures in some cases. It is of interest that about a fifth of the Corps projects authorized in 1962, involving initial gross investment costs of almost a third of the total, have an assumed life of a hundred years, which is the maximum life currently in use. An increase

<sup>12</sup> See Luna B. Leopold and Thomas Maddock, Jr., *The Flood Control Controversy* (Ronald Press, 1954), p. 137, and Otto Eckstein, *op. cit.*, p. 126.

<sup>13</sup> The rate currently used in project planning is 3 percent. This is applied to both non-federal and federal investment.

in the nominal life of a project has a very sizable effect on the B-C ratio if a low rate of discount is in use.<sup>14</sup>

*Evaluation of Demand.* It is admittedly a difficult task to estimate demand functions for many of the services derived from water because of the lack of conventional markets or because of the limited range of price movements. While space does not permit an evaluation of the demand estimates as a whole which appear in the 1962 project reports, it is possible to point out significant errors of concept that are made in estimating some of the demand components. Some of the specific deficiencies may be summarized as follows:

1. Willingness to pay for municipal and industrial water supply is assumed to be measurable by the cost of providing equivalent services by a single purpose alternative. Unfortunately this procedure fails to determine whether buyers are willing to pay this cost for the service or anything at all. If we can be sure that the special circumstances are in fact present which permit this procedure to yield an unbiased result, the demand function can be estimated more directly.

2. Physical standards specified by "experts" are assumed to represent demand for water quality. The monetary dimension of this demand is assumed to be measurable by the cost of achieving the expert's physical standard (e.g., 4 ppm of dissolved oxygen in a stream) by a single purpose alternative method. However, this procedure does not get at the value of incremental improvements in water quality.

3. For recreation, including fish and wildlife benefits, arbitrary prices are used to value the services. The statistical foundation for these prices is not revealed in the reports. The same values are applied to widely differing situations, implying a most unusual set of demand functions.

4. It seems to be generally assumed that demand for municipal and industrial water will increase proportionately or in some simple relation to the increase in population or economic activity. The implicit view seems to be that price elasticity of the demand for water is zero—a view that is demonstrably wrong.

5. In the reports examined no consideration was given to intangible values, such as preservation of scenic areas. Current evaluation standards do require such consideration, however.<sup>15</sup> It will be of interest to observe how this requirement is implemented. For several projects now being planned, consideration is being given to intangible factors.

It is evident that the federal government is moving in the direction of justifying water projects on the basis of benefits that are particularly

<sup>14</sup> An assumed 100-year project life decreases annual costs by 18 percent over a 50-year assumed life where capital costs are 80 percent of total annual costs, a 3 percent discount rate is applied, and the noncapital cost stream is level. With a benefit stream increasing at an absolute rate of 2 percent of the initial year's benefit, the combined effect is to improve the B-C ratio by 35 percent. With an 8 percent discount rate the combined effect is to improve the ratio by only 4 percent.

<sup>15</sup> S. D. 97, 87th Cong., 2d Sess.

difficult to evaluate, such as recreation, municipal and industrial water supply, and water quality improvement. For example, benefits attributed to recreation, water supply, and pollution abatement were 27 percent of the total benefits of flood control projects authorized in 1962 (gross investment, \$2.6 billion) as compared with 3 percent for those authorized in 1950. Substantial advances in procedure will be necessary before monetary values can be assigned to some of these benefits with any assurance. Meanwhile it is doubtful that estimates of these kinds of benefits in monetary terms should be utilized as a basis for project evaluation and design. Existing procedures imply a degree of precision that certainly is not warranted by the facts.

*Exploring the Production Function.* Although estimates of costs for water resources projects as designed may be reasonably reliable, there is good reason to suspect that lower cost alternatives for supplying the same services tend to be neglected. It is quite reasonable to anticipate that an agency responsible for planning water facilities would display an unconscious bias in favor of the kinds of facilities the agency can build and operate. Since the federal water planning agencies are also in fact construction agencies, they tend to favor structural alternatives. They are reinforced in this practice by local beneficiaries, if, as is often the case under present federal policies, the nonstructural alternative is more costly to the local area than the federally subsidized project.

Our review of reports on Corps projects authorized in 1962 suggests that this problem is most serious for flood control and water quality improvement. As others have maintained,<sup>16</sup> nonstructural measures such as regulation of flood plain use may produce greater net benefits through flood damage reduction than structural protection of the flood hazard area. There is little evidence that such alternatives were systematically considered in the design of the flood control projects which we reviewed. It is also significant that existing federal policy tends to foster use of reservoirs for flood control purposes instead of levees and channel improvements because local beneficiaries must pay a substantial share of the cost of providing levees and channel improvements, whereas the federal government pays the full cost of reservoirs.<sup>17</sup>

In the long run this problem may be equally serious for water quality improvement. In the few reports reviewed in which benefits were claimed for water quality improvement, there is no evidence that alter-

<sup>16</sup> Especially Gilbert White in various studies of nonstructural alternatives for reducing flood losses.

<sup>17</sup> A few years ago a team of prominent consulting engineers estimated that flood protection could be provided Kansas City at much lower cost through levees and channel improvement than through reservoirs, but the plan recommended would have entailed large costs to Kansas City. It was never adopted and reservoirs are being built instead. See "Report on Flood Protection in Kansas River Basin," by Abel Wolman, Louis R. Howson, and N.T. Veatch (mimeo., 1963).

native means of achieving equivalent objectives, such as in-plant treatment or infrequent degradation of water quality, were seriously considered. In part because of views expressed in or seemingly implied by the recent Senate Select Committee Report,<sup>18</sup> it has been widely accepted that the avoidance of serious degradation of water quality requires an enormous amount of reservoir storage to increase stream flows during low flow periods. Possibly of greater significance is the fact that Corps spokesmen in projecting their future program assume that beyond secondary treatment a very large increase in reservoir storage to augment low flows will be required for maintenance of water quality.

### *III. Improvement of Efficiency in Satisfying Demands for Water Services*

In spite of the attention that the structure of relationships for water investment decision making has received in the past, more research on how investment decisions actually are made and on how the structure operates is greatly needed. Improvement is difficult, but experience shows clearly that significant, if limited, improvement is possible. There are many participants in this complicated process, and it is possible to change the forces acting on them, to change the standards governing plan formulation and evaluation, to alter operational relationships, and thus change the quality of the resulting investment decisions. The initiative for change can come from any of the participants in the process, but perhaps the most important force for change is the reasonableness of a better application of the principles of economic efficiency.

The desirability of two modifications is almost self-evident, and immediate attention is merited. One of these is to place greater reliance upon charges and prices in the allocation process. It is extremely doubtful that the income redistributive consequences of existing subsidy provisions achieve any clear social objective. If the direct beneficiaries were required to pay for the services they receive, political support for projects would more accurately reflect their social value. Probably no other single measure would contribute more to the attainment of efficiency in satisfying demands for water services and in decisions such as location that are presently distorted by subsidized prices. Such a change may be difficult to bring about for flood control, navigation, and irrigation because existing subsidy policies have become so firmly established, but particular attention should be directed to water quality management which promises to demand such large investments in the future. As has been so well demonstrated in the Ruhr, penalty prices adjusted to the amount and kind of effluent can be used as an effective

<sup>18</sup> *Report of Select Committee on National Water Resources*, Senate Report No. 29, 87th Cong., 1st Sess.

device for stimulating economically efficient behavior where pollution is a serious problem.<sup>19</sup>

The other measure which we believe would foster efficiency is an independent audit of planning activities. Such an auditing unit should be capable of assessing the adequacy of the planning and evaluating procedures. In addition, its examination of project evaluations themselves would involve three kinds of problems. First is the accuracy of the estimates and the data on which they are based. Second is the question of the size of the project; that is, whether marginal net benefits are zero. Third is the all-important question of whether the full range of significantly different ways of satisfying water demands have received adequate consideration. To be effective the group conducting the audit should be independent of the planning and construction agencies. The audit would be made while projects were being planned and not after issuance of the report. The audit unit should not have authority to overrule the planning agency, but through reports available to the public it would illuminate what was actually being done.<sup>20</sup>

Benefit-cost evaluation practices could be improved in several respects. The change that would probably have the most significant effect would be the adoption of a more realistic interest rate. In addition, federal policy might more directly require evaluation of nonstructural alternatives to flood control and non-reservoir alternatives to water quality improvement. If this were done and the evaluations reviewed by an independent audit unit, there would be greater assurance that proposed projects tended to maximize net benefits.

As a final point it would seem desirable not to mislead ourselves by assigning monetary values to benefits when no better values can be developed than are now available for some purposes such as recreation. Instead it would be preferable to develop alternative plans with and without the services in question. The cost of providing the service could then be compared with estimates of the service to be derived expressed in physical terms and a judgment made as to whether the benefits are justified by the cost.

<sup>19</sup> See Allen V. Kneese, "Water Quality Management by Regional Authorities in the Ruhr Area with Special Emphasis on the Role of Cost Assessment," presented at the Regional Science Association Meeting, Pittsburgh, 1962 (to be published).

<sup>20</sup> This is not a new idea although we have defined somewhat differently than others the functions of such a unit. See U. S. Commission on Organization of the Executive Branch of the Government, *Report on the Department of the Interior* (GPO, 1949), pp. 2-6. The Board of Engineers for Rivers and Harbors is, in effect, an audit unit within the Corps of Engineers. As a part of the Corps of Engineers it has, we believe, certain limitations. Nevertheless, our observation indicates that it has a salutary influence.

## THE MARINE FISHERIES: A PROBLEM IN INTERNATIONAL COOPERATION

*By* JAMES CRUTCHFIELD  
*University of Washington*

### I

Any discussion of the economics of the high seas fisheries would appear grossly out of place in meetings geared to the theme of efficiency. Among the resource-oriented industries (a group studded with exceptions to general rules about economic maximization) the fisheries stand out as a most recalcitrant performer.

A few examples will illustrate the point. Throughout the North Pacific we find large fleets of vessels, ranging from one-man gill net boats to elaborate factory ships and mothership units, pursuing salmon which, if left to their own devices, will appear in concentrated schools at the mouths of the rivers of their birth, in prime condition, virtually ready to swim into the processing plant. The vessels would be larger and the industry more capital intensive were it not for the quaint custom of reducing excessive fishing pressure by eliminating the more efficient methods of catching salmon. Until the 1950's fishermen in the highly valuable Bristol Bay red salmon operation were still required to use sailboats in the interest of "conservation," and Alaska still forbids the use of any vessel more than fifty feet in length. In the North Sea thousands of subsidized vessels scour the fishery in the face of convincing evidence that a substantial reduction in fishing effort would bring an increase of perhaps 50 percent in physical yield within a fairly short period. This is the only industry, to my knowledge, ever to be subjected to the counsel of an expert hired by a major government agency to reduce its efficiency in the interest of conservation.

The situation has its ludicrous aspects, but the surrounding circumstances are more sobering than comical. Recent controversies over high seas fisheries have put severe strains on major alliances vital to the political and economic stability of the free world. The United States itself is currently involved in running conflicts over marine fisheries with Japan, Canada, the Soviet Union, and several of the Latin-American countries. Thus far no missiles more deadly than codfish have been launched, but the more insidious device of retaliatory restrictions on international trade is being employed with increasing frequency at the bargaining table.



## II

What is really at stake? On the one hand, there are platitudes in plenty about the boundless wealth of the sea. To the extent that these rest on any scientific basis, they refer to the yield that might be expected if the organic resources of the sea could be harvested at the points in time and space where the total weight is greatest. It might indeed be possible to produce some millions of tons of protein-rich plankton annually but it could hardly be termed food in an economic sense. At the opposite extreme, estimates based on the physical yields to be expected from extension of present techniques on presently productive grounds directed at presently marketable species suggest a maximum annual output less than twice the present level of about 45 million metric tons [1].

Cast in more appropriate economic terminology, the prospects could be summarized as follows. Supply functions based on present techniques would become relatively inelastic at outputs of perhaps 60 to 70 million tons. On the other hand, the fishing industry is in the throes of technological changes that have shifted supply functions rightward by extending both the geographic mobility of fishing gear and the range within which previously unused species are substitutable for those in limited supply. The potential economic yield appears great enough to warrant considerable concern over the consequences of an unrestricted international scramble to expand the shares of individual nations.

The rationale behind the pressure to expand national fishery operations is more evident in terms of differences in the accessibility of marine resources and in the level and composition of consumer demand in various national economies. For some countries (e.g., Norway, Iceland, Japan, and Peru) exports of fishery products bulk large enough in the balance of payments to warrant real concern. In others, including such diverse nations as Sweden, Norway, Japan, China, and many of the underdeveloped nations of Asia, fishery products make up an important part of the total protein intake.

Consumption of fish appears to be highly elastic to income in underdeveloped areas where population is rising rapidly. Though per capita consumption has been stable or declining in most of Western Europe and North America, aggregate world demand may therefore be expected to increase substantially in the foreseeable future. The present emphasis on the development of high-valued species for export (tuna and shrimp, for example) may give way to a much broader basic pattern of exploitation for domestic consumption as the emerging economies develop more adequate physical and organizational arrangements for utilizing fish products efficiently. The matter is far more complex

than a simple matching of available supplies in the sea with areas of critical malnutrition traceable to protein deficiency. Nevertheless, the correspondence is close enough to warrant a serious look at the possibility of devising more adequate solutions to the problems of internationally shared fisheries than are now being considered.

### III

The commercial fisheries of the world rest on the productive capacity of a series of organic populations. Like a forest—perhaps the most similar type of useful living resource—a fish population exists in an enormously complex ecological environment. It is impossible, however, to observe directly the size and composition of a specific fish population, nor can environmental conditions be controlled sufficiently to isolate the impact of specific changes known to affect it. In most cases exploratory fishing is prohibitively costly, and we rarely acquire sufficient statistical data to permit even rough approximations of the production possibilities from a given fishery resource until a well-developed commercial fishery is operating on it.

The key variables determining production possibilities from a fish population can be grouped under four headings: rate of entry into the "fishable" age (recruitment); growth rates of individual fish; natural mortality (from disease, old age, and nonhuman predators); and fishing mortality. In the absence of human intervention, any marine population tends toward a maximum aggregate weight, or biomass, at which net increments to stock from recruitment and growth are exactly offset by decrements from natural mortality. Thus, at zero and at maximum population the instantaneous rate of change in the weight of the fishery population is zero. At intermediate levels, the aggregate weight of the stock, in the absence of other disturbances, will tend to rise toward its maximum value, and the instantaneous rate of change in weight will be positive.

Assuming for the moment that recruitment and growth rates are independent of population size, these relationships can be translated into a simple physical production function. As fishing effort (expressed in terms of standard units) is increased from zero level, sustainable yield—that is, the catch equal to the instantaneous rate of change in biomass in the absence of fishing by man—increases at a decreasing rate, while the number and average size of fish will decline continuously. If the selectivity of the gear with respect to fish of different sizes is held constant, the sustainable yield will peak at some level of fishing effort. Further increases in fishing effort will produce an absolute decline in sustained physical yield. The common sense of this is apparent. Assuming a recruitment rate independent of population and

a sigmoid growth function, fishing by man will yield a larger net physical product as long as the marginal reduction in weight losses from natural mortality is greater than the marginal weight loss resulting from capture of individual fish before they achieve maximum weight.

The assumption that recruitment is independent of population obviously cannot be of completely general validity. For anadromous fish such as salmon, the relationship is critical. Nature is so profligate in her production of fertilized eggs, however, that the case in which the number of fish surviving to catchable size is independent of the total biomass over relevant ranges is the rule rather than the exception. The other assumptions are less tenable. Growth rates are almost certain to be density-dependent, as are some types of natural mortality, and the production possibilities implicit in the foregoing analysis are not necessarily reversible. As the size of the desired stock is reduced through commercial fishing, permanent shifts in predator-prey relations and in relative numbers of competing food users may occur. Moreover, large and frequent shifts in parameters are inevitable in the ecological setting of the sea.

The economist, however, is clearly in no position to point fingers at other disciplines on the grounds that true variables are impounded in *ceteris paribus*. If we are to make any sense at all out of the concept of optimum utilization of a fishery resource, the delineation of physical production functions must be treated as a problem in partial equilibrium, with at least a pious hope that the functions relating the populations in question to their environment may later be quantified. Analysis of the production possibilities of a fishery population is at best a matter of weighing the gains from more informed guesswork against the extraordinarily high marginal cost of adding to basic statistics. The extreme uncertainty about long- and short-run input-output relations colors every aspect of the high seas fisheries.

The superficial similarity of the yield function in a commercial fishery to that of an agricultural operation conceals substantial differences in concept. As pointed out by Gordon [2] the monotonic decline in marginal physical yield is primarily the result of a decline in the size of the "fixed factor" as the level of fishing effort increases. Some elements of Marshallian diminishing returns may also be present; for example, crowding on a particular fishing area may reduce the chance that the gear of each individual unit will contact fish. In general, however, diminishing marginal product is more often a result of a shift in factor proportions resulting from a decline in the number and size of fish in the population exploited than of diminishing returns in the traditional sense.

## IV

To the economist, the analytical definition of production functions and the preliminary quantification of these relationships is but a first step in determining optimal output and factor combination in exploiting the resource. It is slightly unnerving, therefore, to discover that performance of most of the world's fisheries, including the few that are actively managed, has been analyzed purely on the basis of these physical relationships. Maximum sustained yield has been equated with optimum yield, despite the insistence by a few fishery biologists that the latter involves "social" considerations not subsumed in the physical relations. Attention has been focused almost entirely on long-run equilibrium relationships, with no basis for recognition of the economic nature of the decisions to invest and disinvest involved in moving from one point to another on these yield-effort functions. Only a few writers have pointed out the repercussions of changes in the size composition of exploited fishery populations where markets in different countries place different values on smaller and larger fish of the same species or on species produced jointly by the gear employed.

The basic elements involved in conversion of physical to economic magnitudes and the subsequent analysis of individual entrepreneurial decisions in the fisheries have been covered fully elsewhere and need only be summarized at this point [3] [4] [5]. In brief, if a fishery were subject to unified control, the conditions for optimal utilization would differ from those in other resource-oriented firms and industries only in the high degree of uncertainty about short- and long-run input-output relationships.

"Ownership," however, in the conventional sense of the word, is almost never possible in marine fisheries except in the extensive molluscan operations. The physical problem of delineating any useful ownership concept, together with an overlay of institutional arrangements and attitudes with respect to freedom of entry, both national and international, have resulted in common property status for most saltwater fishery resources.

The combination of unrestricted entry to a resource regarded as freely available, a high degree of uncertainty about long- and short-run relations between input and output, and the usual nonsymmetric entry and exit conditions result in a peculiarly unfortunate type of economic behavior in the fisheries. In the absence of specific governmental intervention, any economic rent attributable to the resource will always be dissipated by new entry. Fishing effort is not restricted to the range normally regarded as economically relevant in which marginal physical product is positive. Given favorable price-cost ratios, it is perfectly possible for equilibrium to be reached only at levels of fishing effort

in which marginal physical product is negative—in some cases markedly so. Moreover, this melancholy situation is couched only in terms of long-run equilibrium, in which factor returns just suffice to prevent further entry or exit at a rate of output consistent with stability in the size and composition of the fish stock. The initial development of the fishery, however, is likely to be marked by a surge of new entry that carries capacity well beyond the levels appropriate for long-run equilibrium. It is not only possible but normal for excess capacity to develop quickly and to persist over long periods of time, accentuated by the traditional immobility of labor in the fisheries and the related ability to maintain capital equipment at little or no real cost.

If catch is restricted to levels at which marginal physical product is positive with no restriction on entry, economic rent will still be dissipated completely in the long run regardless of the regulatory method employed. In more cases than not, the prevention of "overfishing" in the physical sense has been achieved by concentrating on control techniques, the direct or indirect effect of which is to reduce the economic efficiency of individual primary producers.

The basic theory of a high sea fishery, whether exploited by single nation or by more than one nation, suggests a bleak economic existence, to say the least. Since economists have not concerned themselves generally with performance of various segments of the fishing industries, the empirical evidence regarding economic performance is not extensive. What it lacks in quantity, however, it makes up in uniformity of findings. Examples include: Pontecorvo's study of the lobster fishery [6]; the work of Donald White and other Boston College economists in the ground fish operations of New England [7] [8]; the Crutchfield-Zellner study of the Pacific halibut industry [5]; the study of the Puget Sound salmon fishery by a University of Washington group [9]; and Sinclair's analysis of the fisheries of British Columbia [10]. In every case the record of economic performance is dismal. Each of these fisheries is characterized by heavy excess capacity, chronically low incomes in areas where labor mobility is restricted, and a distressing lack of internal pressure for innovation or even for adaptation of technical improvements developed elsewhere. In those involving relatively high-valued species—the halibut, salmon, and lobster cases—there is substantial evidence that fishing intensity reached levels at which marginal yields were far into the negative region.

This evidence is supported by the work of several marine biologists. Studies by Beverton and Holt [11] indicate that the yield from the heavily fished North Sea grounds is far below maximum sustainable levels, despite heavy increases in fishing effort in recent decades. Schaefer [12] concludes that even the far-ranging yellow-fin tuna fishery of

the southeastern Pacific may well have reached the point where marginal physical yields are zero or negative; yet new entrants continue to pour into the operation.

## V.

The introduction of multinational fishing effort directed toward a single group of fish populations introduces complications but does not alter the formal microeconomic theory of the fisheries in any significant way. The most obvious complications are: (1) factor prices and factor combinations chosen will vary among nations with different economic structures; and (2) variations in incomes, tastes, and preferences and the availability of substitutes will yield significantly different demand patterns in different countries.

It is evident that differences in consumer preferences and costs make it impossible for two or more nations to reach full agreement on optimal rates and techniques of fishing. The seriousness of the resulting conflicts may be illustrated by reference to the present situation in the northeast Pacific. From the standpoint of the Canadian and American fisheries, optimal utilization would require protection of the limited number of highly valuable species that can be marketed profitably by our high-cost industries, all of which are being exploited at levels very close to maximum sustained physical yield. From the Japanese and Russian standpoint, however, the fishery should be organized around the use of highly efficient gear that can harvest at very low cost a wide variety of species that are perfectly acceptable in Japan's domestic and export markets. As recent headlines make clear, these positions are not always compatible. The situation is complicated further by the importance of fishery exports in the Japanese balance of payments and by the mingling of oceanographic research (some of it of military significance) and fishing operations on Soviet vessels.

Clearly, no single set of decisions as to what fish to harvest, what level of fishing effort to maintain, and what types of gear to employ will be optimal for all four nations. Even if cost and preference patterns were identical, there would remain the question of division of the catch. The situation is not eased by the fact that official representatives of each of the countries involved insist on viewing the problem solely in physical terms, apparently on the ground that some basis for general agreement can be found if all pretend that economic effects are non-existent or unimportant. It is hardly surprising that negotiations among Canada, Japan, and the United States are progressing very slowly; or that the Soviet Union has been—uneasily—excluded from these negotiations.

The frustrating situation in the North Pacific is anything but unique.

The same conflicts exist, in greater or lesser degree, in the North Atlantic, the North Sea, the coastal waters of Korea and China, the Gulf of Guinea, and the Pacific coast of South America, to name only the more prominent cases. The growing mobility of the fleets of the advanced nations and the rapid development of offshore fishing capacity in smaller coastal states promise to extend and generalize the pressure.

Any rational system of exploitation must start with the proposition that a net contribution to economic output can be realized from marine resources only by levying a specific charge to limit inputs; or, alternatively, by limiting entry directly, without attempting to divert rent to a public body through a license fee, tax, or other explicit charge.

Assuming that data are adequate to permit some approximation to rent-maximizing fishing methods and levels, there remain vexing problems of allocating the resulting net economic return. Joint maximization policies are not necessarily preferable, from the standpoint of any single participating country, to those yielding a lower aggregate rent with a greater individual share. The allocation of output among participating nations that would minimize aggregate costs is not likely to be satisfactory to all, particularly those with higher costs. Moreover, the ubiquitous tendency to develop excess capacity in the fisheries suggests that maximization of employment in the national industry may be an important political and economic consideration in some countries. Similarly, participants for whom a particular fishery provides an important source of export earnings might not regard maximization of net economic yield, or even of their own share in it, as the governing objective of international management.

A more serious difficulty in developing a reasonably satisfactory shared fishery arises in the treatment of new entrants. Suboptimal arrangements that might be worked out to encompass the ranges of disagreement outlined above require control over new entry from outside the participating group as well as from within. Yet this may not be feasible, not only on political grounds, but in terms of economic maximization. For example, underdeveloped nations adjacent to major fishing grounds may develop capacity for efficient fishing operations as capital and technical skills accumulate and the effectiveness of internal market organization grows. Their plea for entry into managed fisheries may therefore rest not only on some "right" based on geographical location, or "need" based on nutritional deficiencies, but also on economic efficiency.

If we may indulge in the luxury of contemplating alternatives without reference to their political acceptability, the most intriguing possibility would be some type of supranational fishing entity with exclusive rights over a geographic area that encompasses an appropriate ecologi-

cal unit. There would be no insuperable technical barriers to maximization of rent from the complex of fisheries controlled by such an organization, provided it were free to draw factors from any source and to sell in any market. The allocation of net incomes (and, for that matter, the allocation of employment opportunities) could be handled in a manner which—though something less than an optimum optimum—would still be incomparably better from the standpoint of economic efficiency than any possible arrangement under free fishing. It should also be possible, under unified control of this sort, to deal equitably with emerging states with valid claims to participating rights.

A second proposal (a long step closer to reality) would contemplate regional agreements on shared fisheries, again encompassing an appropriate geographic and ecological area, under which at least two of the most critical problems—determination of the proper level of effort and types of gear to be employed—could be brought under control. In some respects this would be an expansion of the moderately successful arrangements worked out under existing treaties such as the International Convention for Northwest Atlantic fisheries and the treaties covering the Pacific halibut fishery, the Fraser River salmon fishery, the Pacific fur seal operation, and—with serious reservations—the whaling industry. The allocation of rights to participation would be, of necessity, a matter for hard bargaining. But if the only alternative to acceptance of rough-and-ready compromises as to shares in a reasonably efficient fishery is a wide open competitive struggle, it should be possible to reach agreement.

The experience of the United States and Canada in sharing catches from the halibut and salmon fisheries jointly managed by the two countries is illustrative. There is no rational basis, economic or otherwise, for the apportionment scheme adopted in the salmon fishery, and no formal sharing agreement exists in the halibut convention. Nevertheless, both government and industry in each country have been quite willing to accept the resulting apportionment as long as the aggregate catch is protected. The situation would be less manageable, of course, in instances where factor costs and appropriate factor combinations lead to widely differing preferences as to the level of catch and the proper type of gear to be employed, but even an agreement dictated by any one of the participants would be preferable to unrestricted exploitation.

## VI

Space limitations preclude any detailed examination of the legal environment within which internationally shared high seas fisheries must operate. The interested reader will find authoritative discussions in [13] [14] [15] and [16], and an excellent review in Van Cleave



and Johnston [17]. The essentials may be summarized briefly. The historic concept of freedom of the seas is clearly incompatible with any program of rational exploitation of marine resources. To the extent that the doctrine has been applied to fisheries, it rests on one complete misconception and another not far removed from total error. The first assumption is that the resources of the sea are inexhaustible, or at least that man's harvest is so small relative to the total stock that it can exert no appreciable effect on marine populations. The second is that appropriation of marine resources is technically impossible.

The first is demonstrably false and the second holds true only if one rules out the possibility of unified management of marine resources under international agreement. From the standpoint of marine resources, the general principle of freedom of the seas has been successively modified from two directions. First, the elaborate body of law relating to territorial waters and the rights of coastal states sets forth a body of exceptions that seems to grow continuously. For purposes of our discussion, only one point need be noted: neither the issues nor the alternatives offered in discussions of territorial waters imply any recognition of economic maximization, properly defined, as a major objective.

Expansion of territorial waters and of special zones for control over fisheries (or for "conservation") obviously have significant repercussions, particularly with respect to the sharing of catch and aggregate employment opportunities. Quite apart from broader considerations of international welfare, however, the fish themselves seem indisposed to accept such solutions. In few of the critical areas would extension of territorial waters or of special fishery control zones actually exclude participation by other nations. It might well make fishing more expensive, but the ability of highly efficient large-scale fishing units to harvest fish outside the limits of territorial waters is sufficiently great to undermine seriously even the conservation argument for extension of control by a single nation.

Second, both the principle of freedom of the seas and concepts of territorial waters have been modified through bilateral and multilateral treaties, some of which have been successful in preventing serious physical depletion of high-valued marine populations. Their extension to other geographical areas and to other groups of fisheries offers real hope for more rational exploitation of the living resources of the sea.

If they are to accomplish anything significant in that direction, however, the treaty approach to multilateral management must be modified in two major respects. First, it is essential that any fishery treaty consider economic maximization as a significant objective. In virtually

every existing treaty the objectives are specified only in physical terms, usually in one or another variation of "maximum sustained physical yield." Even this unsatisfactory criterion is further modified in some instances to require the use of fish for food even where other usages would yield higher economic returns. The treaty device, appropriately extended, may well be the vehicle through which real progress can be achieved in obtaining maximum human welfare from marine resources; but it most certainly will not serve that purpose until the significance of net economic yield is clearly recognized as an objective and implemented in substantive provisions of the agreement.

The treaty technique must also be modified to permit both restriction of total fishing effort when necessary and the inclusion, on the basis of some acceptable criteria, of new entrants. If the problem of entry is ignored, there is simply no way to develop a continuing net economic return. At the opposite extreme, the American-Canadian principle of "abstention"—the notion that a fully exploited and scientifically managed fishery should thereafter be closed to all others—however desirable from the standpoint of the original participants, cannot hope to win acceptance as a general rule. It would appear that the key to the entry problem may well lie in the matter of employment opportunities. To the extent that fishing nations are not under heavy pressure to utilize excess capacity by expanding the range of their fishing operations, a share in the net rent from a multilateral, controlled fishery might be a perfectly acceptable alternative to direct participation. If restrictions on international trade in fishery products could be reduced, the performance achieved under an ideal supranational fishing enterprise might be approached via the regional treaty technique.

## VII

It was suggested recently in a national news magazine that international arrangements to deal with the investigation and utilization of Antarctica and outer space should look to the law of the sea for guidance. This may offer hope for penguins and whatever astral beings may exist beyond our globe, but it bodes ill for the long-suffering human race.

There is a very real danger that the pressure to resolve increasingly serious conflicts over high seas fisheries will lead to new concepts of international law that do nothing to meet the pressing need for economic rationality in the harvesting of marine resources. Once established, such principles harden into dogma with alarming speed.

Present biological and economic concepts of the theory and practice of fishing could provide a basis for vastly improved economic performance in the high seas fisheries. Yet nothing short of a sweeping revision

of both the objectives and methods of fishery regulation and an equally clear break with the traditional fabric of international law will be required if we are to make more intelligent use of marine resources.

## REFERENCES

1. Food and Agriculture Organization of the United Nations, "International Conference on Fish in Nutrition" (mimeo:) (Rome, 1961), esp. papers by G. L. Kesteven and G. Meseck.
2. H. Scott Gordon, "On a Misinterpretation of the Law of Diminishing Returns in Marshall's Principles," *Canadian J. of Econ. and Polit. Sci.*, Feb., 1952.
3. ———, "The Economic Theory of a Common Property Resource," *J.P.E.*, April, 1954, pp. 124-42.
4. A. D. Scott, "The Fishery: The Objective of Sole Ownership," *J.P.E.*, April, 1955, pp. 116-24.
5. J. A. Crutchfield and Arnold Zellner, *Economic Aspects of the Pacific Halibut Fishery*, Fishery Industrial Research, Vol. 1 (U. S. Department of the Interior, Washington, 1961).
6. Giulio Pontecorvo, "Regulation in the North American Lobster Fishery," in *The Economics of Fishery Regulation* (FAO Fisheries Report No. 5, Rome, 1962).
7. Donald J. White, *The New England Fishing Industry* (Cambridge, 1954).
8. E. J. Lynch, R. M. Doherty, and G. P. Draheim, *The Groundfish Industries of New England and Canada* (U. S. Fish and Wildlife Service Circular 121, Washington, 1961).
9. W. Royce, D. Bevan, J. Crutchfield, G. Paulik, and R. Fletcher, *Salmon Gear Limitation in Northern Washington Waters* (Univ. of Washington Publications in Fisheries, n.s., Vol. II, No. 1, 1963).
10. S. Sinclair, *License Limitation—British Columbia: A Method of Economic Fisheries Management* (Ottawa, 1960).
11. R. J. H. Beverton and S. J. Holt, *On the Dynamics of Exploited Fish Populations* (London, 1957).
12. Inter-American Tropical Tuna Commission, Annual Report, 1961, Appendix A (La Jolla, 1962).
13. F. V. Garcia Amador y Rodriguez, *The Exploitation and Conservation of the Resources of the Sea* (Leyden, 1959).
14. Myres S. McDougal and William U. Burke, *The Public Order of the Oceans: A Contemporary International Law of the Sea* (New Haven, 1962).
15. Shigeru Oda, *International Control of Sea Resources* (Leyden, 1963).
16. W. W. Bishop, Jr., "The 1958 Geneva Convention on Fishing and Conservation of the Living Resources of the High Seas," *Columbia Law Rev.*, 1962, pp. 1206-29.
17. Richard Van Cleve and Ralph W. Johnson, *Management of the High Seas Fisheries of the Northeastern Pacific* (Univ. of Washington Publications in Fisheries, n.s., Vol. II, No. 2, 1963).

## EFFICIENCY OF RESOURCE USE IN CRUDE PETROLEUM: ABSTRACT\*

By M. A. ADELMAN

*Massachusetts Institute of Technology*

Politics, regulation, and monopoly aside, the facts of nature would permit petroleum finding-development-production to be stable under competition. Nearly always, marginal cost is rising and exceeds average cost within any given production unit; always and necessarily for the industry as a whole. There are two occasional destabilizers: unforeseen low-cost discoveries and changing discount rates, but they are minor. Hence the current great instability of the industry, with excess capacity and prices under growing pressure, must be explained, not by nature, but by restraints on the competitive process. The origins and effects of the restraints are very different in the United States and abroad.

In the United States, market-demand proration, sheltered by import controls, is an organized waste or at best work relief—a boondoggle costing \$4 billion annually gross (transfer plus net economic loss)—which is harmful to national security. When additional output is needed, as during the Suez crisis, the state regulatory commissions refuse to permit it, in order to clear out inventories and raise the price. This is no criticism of the commissions, who have only carried out their mandate, and were never asked by Washington for higher output; power without responsibility devolved upon them without notice. Also, because total output is fixed and each well has equal “right” to produce, output per well must be so restricted that today the domestic oil-finding industry is in decline. Possibly the states will reform the system, but not probably. The President can destroy it at any time, and may be forced to do so during some future crisis. For national security, it is necessary but not sufficient to get rid of the present system.

Outside the United States, the market is far more competitive. Prices are sharply down since 1957—crude in the Middle East, by roughly one-third, and products in Europe, by somewhat more. But prices are still far above the cost of finding-developing-producing. Most of the profit is today going to the host governments, whose share is increasing.

Social waste in oil finding and development is small and perhaps negligible outside the United States, but much European and Asiatic coal production is maintained partly to protect local labor and capital,

\* This abstract is reproduced instead of the original paper, which was too long to be included in this volume. The original paper will be printed elsewhere.

partly because of the consensus that prices are headed back up. The author expects them rather to fall. Nuclear energy under known technology seems also wasteful even in most (though not necessarily all) of the high-cost areas, specifically the United States East Coast. Even at present prices, only the quotas on heavy fuel oil, which have no basis in national security, make nuclear power possibly competitive. If oil prices decline toward cost, nuclear power will probably not be competitive for fifteen to twenty years. Whether reserves of oil will remain available past 1975-80 at the present very low supply price, cannot be foretold, nor is it of very great present importance. De-emphasizing known nuclear technology in favor of intensive research on new techniques "would be insurance against the scarcity which may possibly come, not waste of resources adding to the plenty which is here today. But the actual prospect is for more duplication and hence for an increasing oversupply in the world energy market, especially the world oil market, with substantial economic wastes, and with a constant and politically dangerous struggle among companies and governments over markets and between companies and governments over the division of profits. The time of trouble may be, if not prevented, perhaps mitigated. The world oil problem is political before it is economic, but the economist's contribution is not therefore contemptible, any more than the meteorologist advising the commanding general of what he can or cannot do, without presuming to say what he should or should not."

## DISCUSSION

ALFRED E. KAHN: As I read Professor Adelman's paper, it makes three major points: first, that competition is feasible in the oil industry; second, that our domestic regulatory policies have produced monstrous wastes; third, that the wastes caused by such monopoly as persists in foreign oil markets are far less important than those attributable to domestic cartelization. He also makes certain predictions—which I have less competence to judge but which strike me as entirely reasonable—about the future course of world oil prices and prospective competitiveness of atomic power.

At the risk of shocking him and myself, let me begin by pointing out that I agree in all fundamental respects with the three major conclusions and their supporting argument. But I have enough reservations, I think, to make productive use of my ten minutes.

With respect to the feasibility of unregulated competition in oil, I must admit to more uncertainty than Adelman. Specifically, I am not satisfied he has given adequate weight to his "occasional destabilizers," and the pressures they may pit against "the strong forces pulling . . . [the industry] toward an equilibrium whenever disturbed."

His argument here is based essentially on the industry's inherent tendency to statically diminishing returns. I would emphasize more than he the additional potentially stabilizing influence—provided, of course, reservoirs were unitized—of user costs, as Paul Davidson did in the March, 1963, *American Economic Review*; but there is no major difference between us here.

But I do not see how one can be quite so certain that the stabilizing tendency of statically increasing costs will in an unregulated market compensate with tolerable efficiency for fluctuations in actual costs and supply over time—a consequence not merely of the accident of discovery or of changing market rates of discount but also of changing market expectations about the future price of oil, technological progress, and, very importantly, the opening up of new areas to exploitation, the latter in turn the result of technological change, possibly sudden political developments in oil-rich countries, and the process of economic growth itself.

Consider, for example, the situation in the Middle East. Adelman's own paper suggests—what I too believe to be the case—that output in this area could at any given time in the last twenty years have been several times as great as it was, at no higher unit cost and of course at a cost far below the American. Admittedly this was a consequence of monopoly and could not have persisted under more purely competitive conditions. But to recognize that the vast expansibility of output in that area, and in North Africa today, at constant cost, could only reflect impurity of competition is not to prove that a more competitive market could have adjusted with tolerable speed to the original appearance of these possibilities.

By all means let us give competitive forces freer play. It is inconceivable to me that their results could be as bad as what we now suffer. But let us remember that competitive markets, too, can be intolerably and wastefully

unstable, and there are special reasons why this might be so in oil. Statically increasing costs are no insurance against the cobweb phenomenon—and there are evidences of this in oil—or the explosive effect of changing expectations.

Still, I cannot in the end disagree with Adelman's appraisal of our domestic regulatory policies. I simply cannot myself find in the possible destructiveness and instability of a more purely competitive regime any defense for the enormous wastes inherent in the system of fair shares. My differences with his argument on this second point would all be on matters of detail, but I think they are worth listing: (1) The reader might not recognize, from this paper alone, how much of an improvement market demand prorationing was over the uncontrolled law of capture. I think, too, that Adelman underrates the promise of the tendency of state regulatory agencies to base production quotas increasingly on an acreage rather than on a per well basis, as industry leaders have themselves become painfully cognizant and increasingly vocal about the costs of the present system. (2) I am not sure what Adelman means when he ridicules the view that "prices are not made by supply and demand, but by goodies and baddies, the latter . . . through administered prices." If the reader infers from this that prices in this industry are determined by impersonal, competitive market forces and not at all by the policies of regulatory agencies or company officials, he will be mistaken. (3) Adelman errs, I think, in stating flatly that prorationing is a "menace" to national security. The combination of production control and the various tax privileges accorded this industry have given rise to a huge idle production capacity that could be activated at once in the event of emergency. The contribution of this excess capacity to national security is by no means unmitigated, let alone costless; but it cannot be ignored. (4) I think it necessary to point out, finally, that the \$3.6 billions, to which Adelman refers as a "rather conservative estimate" of the "annual charge of this organized waste" must not be misread as representing the real cost of prorationing to the American economy. The figure comes from multiplying our annual domestic consumption by the \$1.00 a barrel by which it is estimated the price of oil would fall if import as well as domestic production controls were dismantled. For various reasons it grossly exaggerates the real costs of prorationing alone.

On Adelman's third major point, the contrast he draws between the circumstances at home and abroad is instructive. But in pressing the contrast, it seems to me he has at some points minimized unduly the persistence of monopoly in foreign production. The evidence of how much remains is the fact, emphasized several times by him, "that the price, though much reduced, is still far above finding-developing-operating costs." This fact, and the presence of a great "surplus . . . of available reserves" that "can be turned quickly into capacity by a relatively small development investment . . ." are both explainable only in terms of severe impurities of competition.

So it appears the world industry, too, and not just the American, is playing the game of "fair shares," and hence in effect "deliberately holding back low-cost supplies to make room for the higher" (which Adelman denies). The explanation must be a continuing private recognition of oligopolistic interdependence. And its effect, like that of domestic cartelization, but by no

means so wastefully, is to encourage overinvestment and a growing excess capacity.

None of this, I repeat, is to quarrel with the basic point that in this industry the wastes of compulsory cartelization with free entry surely far exceed those of monopoly.

DONALD J. WHITE: Will the day ever come when West Europeans will let the Icelanders catch their fish for them? Is there any hope that economic efficiency considerations—the maximization of net economic product rather than some other criterion such as attainment of maximum sustained physical yield—may come to receive more attention and bear more weight in fisheries policy making?

Professor Crutchfield's comprehensive yet penetrating paper calls a spade a spade. Given the "extreme uncertainty about long- and short-run input-output relations [which] colors every aspect of the high seas fisheries"; given the consideration that "in the absence of specific governmental intervention, any economic rent attributable to the resource will always be dissipated by new entry"—even to the point that equilibrium will be reached only at levels of fishing effort in which marginal physical product is negative; and given the fact that it is not only possible but normal for excess capacity "to develop quickly and to persist over long periods of time," then surely, as Professor Crutchfield says, "the basic theory of a high seas fishery, whether exploited by a single nation or more than one nation, suggests a bleak economic existence to say the least." When international differences in cost and market structures and in economic and political goals are taken into account and when such formidable myths as the alleged inexhaustibility of the resources of the sea are acknowledged, the situation on balance tends to become even more inimical to the achievement of economically efficient exploitation of fishery resources.

Where such dismal equilibria seem preordained if nations permit their fishermen simply to do "what comes naturally" under completely free entry conditions and even subsidize their fishermen's efforts when the only result must be even less economic efficiency, why have not the nations and their fishermen been more interested in negotiating more economically viable arrangements and regulations?

Professor Crutchfield gives quite well the full litany of responses to this question. Perhaps more important than any other, however, is the response that economic analysis and empirical work and education with respect to existing findings have not gone far enough or been sufficiently convincing and understood by the parties to generate leverage in favor of seeking arrangements other than destructive competition so that more nearly optimal economic results might be obtained. The economists' most important task, then, is to raise the level of economic understanding, for until more progress is made on this, little can be accomplished in the way of constructive change.

To build up knowledge and education and improve prospects for fruitful cooperation and policy making, I would supplement Professor Crutchfield's presentation by suggesting that the United States take the lead in proposing



an International Fisheries Year, analogous to the International Geophysical Year, with all relevant disciplines to be brought into play. This is not meant to be a "gimmick." It is meant to spotlight the crucial role of research and education, to provide more momentum for continuing multidisciplinary research and increased international understanding. The suborganization of the arrangement could be structured to reflect regional or area interests. The effort would be analogous to current labor and management efforts in the United States in a variety of industries to make possible through joint study on a continuing basis more knowledgeable collective bargaining and to minimize discord.

During all this, the appropriate parties could pursue attempts to reach better regional accommodations. As Professor Crutchfield has noted, for example, the United States and Canada have successfully shared catches from the halibut and salmon fisheries jointly managed by the two countries. That the countries have operated on biological optima rather than economic optima and hence have tolerated wasteful economic arrangements while achieving some measure of improved physical productivity, is in my view just more evidence of the need for education and for stepping up the dialogue between economists and biologists in particular, so that the two professions will become more fully effective partners in giving industry and other policy-makers the knowledge so vital to constructive negotiations.

In suggesting the International Fisheries Year, I do not mean to imply that international meetings have not been held. They have occurred from time to time on a limited conference basis on one phase or another of the industry's problems. But the current undertakings or plans to expand fishing effort in practically all sections of the world lend emphasis to the need for dramatic action to make sure that knowledge and understanding have the fullest opportunity to be brought to bear to prevent sheer waste and insure minimum friction.

Finally, it might be noted that the development of shortages of fishermen in advanced industrial areas, notably now in Germany, for example, may ease the problem of reducing fishing effort. This consideration is of course extrinsic to the fisheries management problem per se, and this probably accounts for Professor Crutchfield's failure to allude to it. But fishing is hard physical work, risky, uncertain, and involves long hours at sea away from home. It should not be surprising, therefore, if rising living standards, provided there is reasonably full employment, tend to lead to a substantial reduction in the supply of fishermen, with favorable efficiency consequences.

NATHANIEL WOLLMAN: When dealing with water resources the economist faces a task of more than ordinary interest because he is confronted with a commodity that is just now in transition from a free to scarce good. Presumably the economist and political scientists ought to be able to wield their instruments on water with the same facility that they can upon other scarce resources. But in the present state of knowledge their technical skills outrun their material, and it will take some further probing before the two are in phase. The heavy reliance upon inference in the Fox-Herfindahl paper illuminates our present position.

The authors have prepared a lucid and persuasive paper. The assumptions are reasonable and the logic is persuasive. However, when one studies their paper carefully it is surprising to see how little evidence is provided to support their argument. In part this is the result of brevity requirements imposed by the program chairman, but it is also the consequence of the present state of the art.

They do not assert that marginal equality between benefits and costs is violated, but they raise the suspicion that it may be. In support of their suspicion they offer one case out of the 178 they examined. I am free to conclude, therefore, that in 177 cases the marginal principle was adhered to, which seems pretty good.

They say that if a higher interest rate were used in computing benefits and costs, from 9 percent to 80 percent of the projects examined would have benefit-cost ratios less than unity, but they do not suggest how high the interest rate should be nor do they offer any evidence that it is now too low, except reference to what is avowedly controversial literature.

They object to the implicit assumption that the price elasticity of the demand for water is zero. They object to the assumption of constant technical coefficients and to the failure of construction agencies to take all technical substitutes into account in constructing a program. But all of their argument is inferential rather than based upon direct evidence. They fail to cite any of the 178 cases to demonstrate that the facts of demand are different from the assumptions of demand. Without facts to prove the contrary, it is perfectly reasonable to assume a very low, perhaps zero, price elasticity. For example, if I anticipate that other regions are going to encounter water restrictions and I know that the costs of water are low relative to other costs—both assumptions being plausible and supported by evidence—I can conclude that for any given region the price elasticity of demand will be low over a relatively wide range of price variation.

The authors make four major recommendations designed to increase efficiency of resource allocation: (1) Put more reliance on market pricing. (2) Use a higher interest rate in discounting benefits and computing costs. (3) Create an independent audit unit. (4) Substitute common sense for dubious monetary values of intangibles and nonvendibles. My general reaction is that these may be desirable improvements but they are not supported by the body of analysis.

Let me illustrate. The authors' most emphatic recommendation is to make greater use of charges and prices. "Probably no other single measure would contribute more to the attainment of efficiency," they assert. However, their paper inadequately supports this proposition. Even if one grants that the present decision-making process is loaded in favor of benefits and against costs, one cannot logically jump to the conclusion that market pricing yields superior results. There is a missing link: an appraisal of the comparative efficiency of market pricing relative to the techniques now in use. This link has not been supplied.

One might argue that water markets are inherently imperfect because of limited mobility of supply, extreme interdependence among users, the need for heavy capital investment, and uncertainty of physical supply. The authors

themselves point out that the market for intangibles is too primitive to yield satisfactory information, yet is increasing in relative importance. If one were to rely upon inference from first principles as do the authors, one could just as easily go in the opposite direction and conclude that greater efficiency in water use requires more rather than less administrative control. Since they wish to assign more responsibility to the market process for establishing the standard of efficiency, they are obligated to say how far they intend to use the market, when they will accept market solutions, and when they will reject market solutions.

I can agree with their suggestion of an independent audit unit because some experimentation is in order, but why didn't they recommend a basin commission or interagency commission? Certainly the addition of a powerless audit unit does nothing to solve some of the deficiencies of our present system as noted by one of the coauthors recently in another paper.<sup>1</sup> The deficiencies which he specified were three in number: "(1) costly and time consuming nature of the planning effort; (2) lack of continuity of the planning process; (3) lack of follow-through from planning to implementation."

I would guess that the independent audit unit would add to the "costly and time consuming nature of the planning effort" without subtracting from the remaining two deficiencies. I wonder why the authors did not propose an agency like the Delaware River Basin Commission, which would have continuity and supervisory power. Instead of an independent audit unit, which would duplicate the scientific task of data collection and evaluation now performed by the agencies, why not have an independent Board of Analysis and Estimates to prepare the economic and engineering reports that are required. This board would exist independently of the construction agencies, would work for all river basin commissions, but be identified with none of them. The existing agencies would still be responsible for fulfilling their missions but only within the limits of the plans prepared by the basin commission.

I agree with their last recommendation regarding intangibles—but I wonder why they suggest that a higher discount rate be used in computing the benefit-cost ratio, when the same problem of valuation that applies to intangibles applies to selection of the interest rate.

Where does this leave us? I think that the Fox-Herfindahl paper provides a starting point for two parallel lines of activity. Since investigation of the Corps' projects did not compel the authors to discard the hypotheses that inefficiencies exist, the next step would be to provide positive proof and to disclose how serious and with what frequency the inefficiencies are encountered.

The other line of activity would be research designed to reveal just how far market processes should be substituted for administrative action in order to raise the level of economic efficiency. The research should also be directed toward the opposite question as well; namely, where should administrative action take the place of market processes and to what extent should market processes be used as instruments of administrative decisions.

<sup>1</sup> Irving K. Fox, "Water Resources Planning: An Historical Review and an Examination of Recent Experience" (Nov. 19, 1963, mimeo.).

# EFFICIENCY IN THE GOVERNMENT SECTOR

## FISCAL INSTITUTIONS AND EFFICIENCY IN COLLECTIVE OUTLAY\*

By JAMES M. BUCHANAN ✓  
*University of Virginia*

Fiscal institutions have been critically analyzed for their effects on the "efficiency" of the private sector but almost no attention has been given to the effects of these institutions on the "efficiency" with which resources are allocated to the public sector of the economy. Until quite recently, and with Knut Wicksell as a notable exception, economists have not attempted to analyze political or group decisions with the tools at their command. They have neglected the fact that individuals also "behave" as they participate in collective decision processes and that such behavior can be analyzed in terms of a theory of individual choice.

Limitations of space prevent extensive discussion here of an analytical framework as well as the examination of a number of alternative models. Both the type of questions that should be raised and the type of analysis that may be attempted in answering them may, perhaps, be suggested through concentration on a single example. What effects would a change in the effective progressivity of the federal revenue structure exert on the relative size of the public sector? And would such a change tend to produce more or less efficiency in the allocation of resources to the supply of collective goods and services? For simplicity, the analysis is limited to a reduction in progressivity.

### I

Before attempting to answer these questions, it is necessary to ask and to answer a more simple one. How much will a political community choose to devote to the supply of collective goods and services? The outcome will, in general, depend both upon the rules for the making of political decisions and upon the tax institutions that are to be employed to finance the public goods. My approach is based on the presumption that both the political decision rule and the tax institution exist independently of possible budgetary size or composition.

I shall first introduce a world-of-equals model. All persons in the

\* Some of the background work on this paper has been done in connection with a more comprehensive project on fiscal institutions that is supported by the National Committee on Public Finance, under the auspices of the Brookings Institution.

political community are identical in all respects, including the evaluation of the single public good, which, for our purposes, we assume to be purely collective in the Samuelson sense. How much will the group decide to spend on supplying this good collectively? This model is interesting because it is the only one in which neither the political decision rule nor the tax institution exerts an influence on the final outcome, provided only that the tax is a general one that exhibits what has been called horizontal equity. In addition, the outcome will satisfy the necessary conditions for Pareto optimality, second-best considerations aside.

In this model, any general tax will impose the same tax price per unit of public good on each individual. For example, individual tax liabilities will be identical under capitation taxes, proportional income taxes, progressive income taxes, or expenditure taxes. Each person would choose, were he given the power to decide for the community, that the collectivity finance the same quantity of the public good. Therefore, the delegation of decision-making power to a single person (any person), simple majority voting, or unanimity will produce identical and "optimal" results.

## II

A variation on the world-of-equals model allows the analysis to be substantially extended. Assume now that individuals are identical in preferences but that incomes differ. This may be called an "equal-preference" model, and in the familiar geometrical analysis it means only that the preference maps for the various members of the political group are equivalent. Individual evaluations of the collective good now differ to the extent that they are affected by incomes.

What particular characteristics of the preference pattern must be present if a system of proportional income taxation is to satisfy the requirements for "full neutrality"? By full neutrality, I refer to that position where all of the necessary marginal conditions for Pareto optimality are satisfied and from which no political decision rule would generate a change. The set of positions satisfying the full neutrality requirement is a subset of those positions that may be classified as Pareto optimal. In full neutrality the marginal tax-price confronted by each individual is equal to the marginal evaluation that he places on the public good, and the summed tax-prices equal marginal cost. Given these conditions, no individual desires to modify the quantity of the public good that is supplied; all individuals prefer that same quantity. Hence, the political decision rule is immaterial.

Figure 1 illustrates the analysis with an indifference curve-budget line configuration. Income (private goods) is measured along the ordinate; units of collective good along the abscissa. The fan-like array of

budget lines intersecting on the abscissa in  $G$  represents a proportional tax structure. An individual with income,  $Y_1$ , confronts a budget line shown as,  $T_1$ , with the tax-price per unit of the public good indicated by the slope of this line, and, similarly, for all other individuals in the group. For simplicity, we assume that the supply price of the public good is constant over quantity, and that, regardless of the tax institution, each individual faces a constant tax-price per unit of the public good.<sup>1</sup>

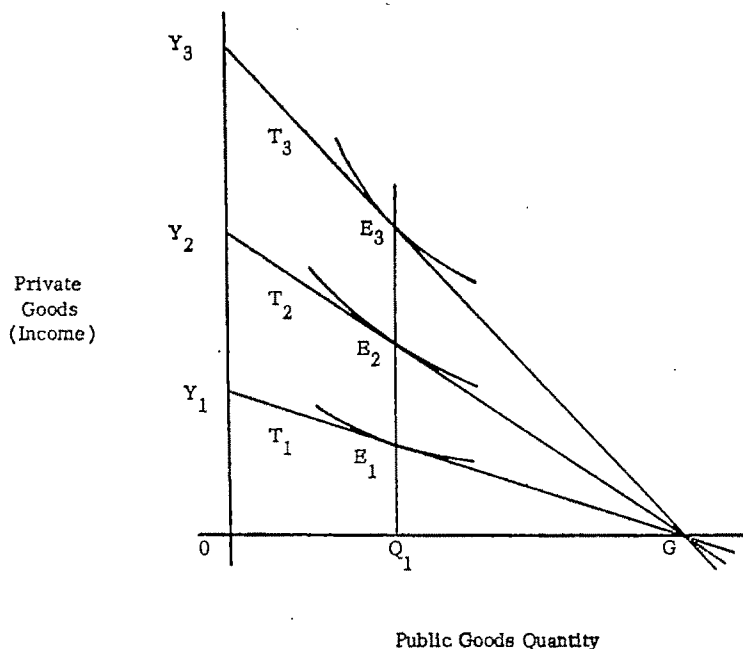


FIGURE 1

Assume now that a specific quantity of the public good, say,  $Q_1$ , is financed by the schedule of tax-prices shown. The preference map configuration that will satisfy full neutrality is obvious; the successive indifference curves must be tangent to the budget lines along the vertical drawn from,  $Q_1$ , as Figure 1 illustrates. The construction could be modified to fit any postulated tax institution.

A more general statement of the necessary condition is as follows: The income elasticity of demand for the public good divided by the price elasticity of demand must be equal to, and opposed in sign to, the income elasticity of the tax-price schedule.

<sup>1</sup> For a similar construction in which tax-price is allowed to vary, see, R. A. Musgrave, *The Theory of Public Finance* (McGraw-Hill, 1959), p. 122. Note, however, that Musgrave's usage of the construction is different from that of this paper.

Income elasticity of demand  
for public good

$$\frac{\text{Income elasticity of demand for public good}}{\text{Price elasticity of demand for public good}} = (-) \text{Income elasticity of tax-price schedule.}$$

Full neutrality is present when this condition is met throughout the range of possible incomes. Under proportional income taxation, the income elasticity of the tax-price schedule is unitary. Therefore, the income elasticity of demand must be positive and equal to the price elasticity in absolute value. Suppose, for example, that the income elasticity over the relevant range is positive with a coefficient of 1.5. Clearly, the price elasticity of demand would have to be negative and also with a coefficient of 1.5 in order that the individual at the two income levels (or, in our model, two individuals at the different income levels) prefer the same quantity of the public good.

If income elasticity of demand is positive and exceeds the price elasticity in absolute value, the wholly neutral tax-price structure must be progressive. If income elasticity is positive but less than the price elasticity, the fully neutral structure of tax-prices must be regressive. And, of course, if the income elasticity of demand should be negative, the tax-price confronted by the individual would have to decrease, not increase, as income rises in order for full neutrality to hold.

### III

The analysis remains severely limited unless departures from the conditions required for full neutrality can be incorporated into it. Figure 2 allows this extension. On the abscissa is measured income; on the ordinate is measured tax-price and marginal evaluation. A proportional income tax is represented by the line  $t_1$ . Three different progressive rate structures are shown; these are  $t'_1$ ,  $t_0$ , and  $t_2$ .

For each given level of public good, a marginal evaluation schedule may be derived which relates individual marginal evaluation to the level of income (private goods). Geometrically, this schedule is derived by plotting the slopes of the successive indifference curves as they cut the vertical drawn from the given  $Q$  in Figure 1. Let us assume, initially, that full neutrality is present under the proportional income taxation schedule,  $t_1$ . In this case, the marginal evaluation schedule, given the quantity of public good,  $Q_1$ , lies along  $t_1$  and coincides with it. This is noted as  $v_1$ , in Figure 2.

Now suppose that a progressive income tax is to be substituted for the proportional tax. Since full neutrality is to be violated, the rule for making political decision now becomes important in determining the outcome. Let us assume, therefore, that collective decisions are to be made under a rule of simple majority voting. It is necessary to distin-

guish two types of departure from full neutrality, which I shall call here symmetrical and nonsymmetrical. Symmetry is defined with reference to the position of the median voter-taxpayer. Under simple majority voting, with single-peaked preferences, the median voter be-

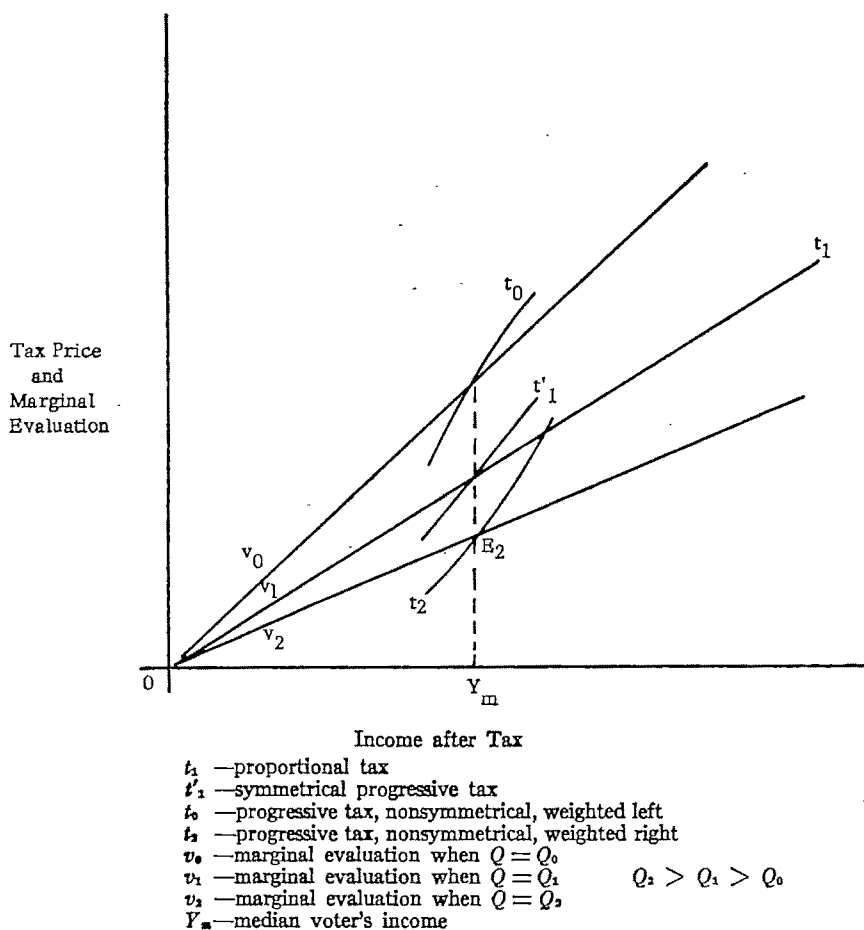


FIGURE 2

comes controlling, and the direction of change in collective outcomes can be predicted on the basis of his behavior.<sup>2</sup> By a symmetrical shift from proportion to progression in rate structure, I refer to a shift that does not modify the tax-price per unit of the public good that confronts the median voter-taxpayer. In other words, symmetrical progression may be introduced by imposing a higher tax-price on the "rich" which

<sup>2</sup>The basic analysis of single-peaked preferences under majority voting is contained in, Duncan Black, *The Theory of Committees and Elections* (Cambridge Univ. Press, 1958).



is precisely offset by lowering the tax-price to the "poor," without changing the position of the middle man. Since his position is not modified by the change, the median voter will not desire to vote any differently. And, since his vote is decisive in this model, there will be no change in the quantity of the public good that is supplied.

Such a symmetrical change is shown in the progressive rate schedule,  $t'_1$ , in Figure 2. The poor will, because of the lowered tax-price, seek more actively to expand public goods supply. The rich, for the opposed reasons, will seek to contract it. But their efforts will be to no avail in the simple majority voting system. The outcome remains unchanged and it remains Pareto optimal since the excess benefits enjoyed by the poor just offset the excess taxes paid by the rich. No bribery or scheme of compensation could be worked out so as to change the allocative results. In effect, the introduction of symmetrical progression amounts to a shifting from one position to another on the conceptual Pareto welfare surface and is equivalent to a set of lump-sum transfers among income groups.

A nonsymmetrical shift is defined as one that does modify the tax-price to the median voter-taxpayer. Such shifts can be weighted in either of the two directions. A shift from proportion to progression can increase or decrease the tax-price that must be paid by the median voter. If the change is concentrated on reducing the tax-price to the lowest incomes groups, for example, by simply raising the personal exemption while leaving the rate structure above the exemption proportional, the tax-price to the median voter will be increased. If this is the case, which we can call a weighting to the left, the median voter will choose a smaller quantity of the public good, and because his vote is controlling in any majority coalition, the collective result will embody less public spending. One such system is shown as  $t_0$  in Figure 2. If, on the other hand, the change is concentrated on increasing the tax-prices of the highest income groups, for example, by fully taxing realized capital gains, the tax-price confronted by the median voter will be decreased, the curve,  $t_2$ , illustrates this case.<sup>3</sup> There will result here a somewhat larger collective outlay on the public good.

In either case, a nonsymmetrical change will modify the final outcome, and the necessary conditions for Pareto optimality will be violated, assuming that they were satisfied before the change.

#### IV

There is, of course, no way of knowing to what extent the necessary conditions for Pareto optimality are satisfied at any particular point in

<sup>3</sup> Except under certain elasticity conditions, the income of the median voter, after tax, will be changed by any nonsymmetrical shift in the tax structure. This complication has been ignored in drawing Figure 2 since it does not affect the analysis.

time. The analysis does enable us to predict, however, the direction of change in total outlay quite independently of the knowledge that we may have concerning individual marginal evaluations. Given the prevailing distribution of personal income, it seems highly probable that, in the neighborhood of the median voter, the federal tax structure takes the general shape shown by  $t_2$  in Figure 1. If this is correct, and if an overall reduction in the progressivity of the system tends to shift the rate structure in the direction of that shown by either  $t_1$ ,  $t'_1$ , or  $t_0$ , in the neighborhood of the median, then the tax-price faced by the median voter-taxpayer will be increased. As a result, he will tend to choose a smaller outlay on the public good than under the more progressive structure. In other words, I am suggesting that the predicted reduction in progressivity will itself be nonsymmetrical in a specific direction.

We know that the progression in the federal income tax becomes significant only over income ranges that exceed the income level of the median taxpayer. This fact insures that, in more general terms, political equilibrium is represented by a position similar to  $E_2$  in Figure 2. If, from such a position, the tax-prices imposed on high-income taxpayers are reduced more, in absolute value, than the tax-prices on low-income taxpayers are increased, the tax-price confronted by the median taxpayer must be increased. This remains true despite "tax reductions" that extend over the whole income scale. The relative change in the tax-price at which public goods may be "purchased" through the fiscal process is the determining influence on the behavior of the individual voter. And if this is increased, he will choose to consume fewer public goods than he would have chosen under an unchanged tax structure.

This hypothesis seems intuitively plausible. To an extent, at least, the median or middle-range voters support the current level of public spending for the by-product distributional effects that can only be secured via the fisc. A reduction in progressivity can only decrease these redistributive benefits. And the middle-range voters can be expected to react against the increases in the level of tax-price that they confront. }

It is considerably more difficult to predict the effects of the reduction of progressivity on overall efficiency in the organization of the public sector than it is upon the direction of change in the size of this sector. To do this, it is necessary to know something about the actual income and price elasticity of demand for public goods. In drawing Figure 2, we have assumed that the income elasticity and the price elasticity are just offsetting over the whole of the income range. (The  $v$ -curves are linear as drawn.) This need not, of course, be at all "realistic," and either "progressive" or "regressive"  $v$ -curves may be more in accord with the data. If, however, the  $v$ -curves in Figure 2 are accepted as reasonable (and there exist certain data in support of this), then the

reduction in progressivity, carrying with it some reduction in overall public outlay, will tend to produce greater efficiency. This is because, before the change, the public sector tends to be expanded beyond Pareto-optimal limits. This, again, seems intuitively plausible. Insofar as the distributional by-product is the basis for the spending on public goods, there will tend to be some expansion of the public sector beyond the optimal allocative limits. On the other hand, should the appropriate *v*-curves exhibit progression, the current rate structure may generate roughly optimal outlay, and any reduction in this outlay would tend to reduce overall efficiency.

I am hopeful that studies currently in progress will reveal something more about the actual shapes of the *v*-curves and the *t*-curves. Only when these studies are complete will definitive conclusions concerning the effects on efficiency be possible. As of now, I am prepared only to suggest, without defending my position too strongly, that the reduction in overall outlay that the reduction in progression seems likely to generate will probably increase rather than decrease the efficiency with which the total resources of the economy are divided between the provision of private and public goods.

## V

The analysis has been confined to what I have called an equal-preference model. The general appropriateness of this model may, of course, be questioned. Surely, income effects are not the only explanation for differences in individual evaluations of public goods. Tastes differ for public goods as well as for private goods. Some people do not like garlic; others do. Similarly, for foreign aid. These specific differences in tastes, however, are more important in relation to single private goods or single public goods than they are when we come to consider the whole package of private goods or the whole package of public goods, or, alternatively, a single public good to be taken as representative of the whole package. It seems plausible to suggest that, in this case, income differences do loom relatively large in explaining differences in the evaluations placed on public goods by individuals.

For the model to be relevant at all it is necessary to assume that the public sector includes only goods and services that provide general benefits to the whole of the population. Insofar as discriminatory benefits are provided to particular groups, with remaining groups wholly or largely excluded from enjoying them, the analysis is not applicable. However, in such cases the familiar institutions of taxation are not applicable either. Implicit in the development of the basic institutions of general taxation there seems to have been the idea that income and wealth differences do reflect, to an extent, differences in willingness as

well as differences in ability to pay. If these is, in fact, no relationship between income and wealth levels and the evaluations that individuals place on public goods, almost all of the institutions of general taxation must produce serious distortion in the allocation of economic resources.

Finally, the model of equal preference may be defended as being relevant on quite different grounds. It has been convenient to present the analysis in terms of a single-preference map. When we come to examine empirical data, however, we must, at best, draw on cross-section studies that report on the behavior of a number of individuals. In effect, our empirical limitations force us, whether we like it or not, to talk in terms of some composite or representative preference pattern. This serves, of course, to rescue our whole model from the apparently restrictive limbo that may have seemed its fate. The elasticities to which we have referred in the criterion for full neutrality are derived from data drawn from the behavior of the whole population, not from the behavior of single individuals.

## VI

I am prepared to admit that other factors, not accounted for in this analysis, may serve to offset these developed here, and, in so doing, to refute the underlying hypotheses that I have advanced. If in reducing the overall progressivity of the federal tax structure the tax-price to the median voter should be increased, as I have suggested, but simultaneously, his tax awareness or tax consciousness should be reduced, the size of the public economy may be expanded and not reduced. This possibility should not be ruled out of account, especially since the reduction in overall progressivity may be accompanied by an increase in the share of indirect taxes (including inflation) in total revenue production. I have also assumed throughout the analysis that the simple models of majority voting retain some relevance for real-world political decision making. This need not be the case, at least in the simple form that the models have assumed here. If, in fact, it should be found that the effective political power of individuals is directly related to income levels, the overall reduction in progressivity may generate an expansion in public outlay from this factor alone. I have also limited the analysis to models of democratic choice. The power of "leaders" in determining the pattern of resource outlay between the public and private sectors has been completely ignored. Insofar as individual citizens do not, ultimately, determine the size of the public sector, the discussion is irrelevant. To analyze the alternative here, we require models of *dirigisme*, not democracy. And this task I leave to those of my disciplinary colleagues who would presume both to advise the despots and to explain their behavior.

## THE STRUCTURE OF GOVERNMENT AND PUBLIC INVESTMENT

By JULIUS MARGOLIS  
*University of California, Berkeley*

This paper will discuss how public investment is affected by the characteristics of the levels of government. I have restricted myself to a few issues in the analysis of public investment which have been of concern to analysts in recent years. There are many features of a structure of governments which are relevant but ignored in this brief paper.

The recent studies in the theory of public expenditures and analysis of public investment have assumed a unitary government. The typical model of the theory of public expenditures is one of decision making via referenda or town-hall meetings—a village democracy. The contrasting model of the analyst of public investment is a government of administrators who measure a stream of repercussions extending indefinitely over time and space and evaluate them with welfare criteria applicable to a national government. In fact, what there actually is, is a structure of governments of differing powers and responsibilities.

A consideration of the structure of governments gives a new perspective to old questions. We might ask whether some of the insoluble problems posed in the theory of public expenditures are worked out through the behavior of the structure. That is, does the structure have some of the characteristics of an industry and market, so that there is an interaction among governments which leads to desirable results.

In addition to neglecting the structure of governments, analysts have tended to collapse excessively the internal organization of government. For those concerned with normative theory of public expenditures, the government is the market place where allocations are made and costs exacted. In some cases, the allocations are the result of an exchange model reminiscent of the perfectly competitive market, while for others it is the battlefield of groups playing all the tricks of oligopolists who are expected to collude. For those concerned with planning criteria, the excitement of conflict is removed and the government becomes a bloodless computer. The government of administrators assumes a welfare function; it estimates production possibilities, and then it optimizes.

The government is more than a market place or a planning office. It is also a source of power and income. Just as the market can be rigged, the government can be manipulated to protect private interests of some constituents. Just as promoters can orient and stimulate the market,

there is a government bureaucracy which can gain from government activities. All three of these views of government—market place, detached administrators, and self-oriented personnel—have some explanatory value for government behavior.

It is clear that public services can be supplied by any of many types of governments. The selection of an appropriate government will be affected by ease of carrying an issue and not simply by the relative efficiency of a centralized or decentralized provision of services. We shall analyze two of the factors leading to preferences for central governments—one which is related to the pattern of decision making and a second which is related to real differences among governments.

Let us consider an abstracted investment process for the local community. Proposals are usually brought forth by the administration or hired special consultants. The proposals are subjected to scrutiny by an elected council or board, an *ad hoc* citizen's commission, outside experts, and finally the citizens are asked to endorse the proposal by supporting or rejecting the bond issue authorization necessary to finance the investment. The requirements for support of the council or electorate may be for two-thirds or some other fractional agreement rather than simple majority. Often other governments may be involved if their grants of support are to help finance, or if there are regulatory agencies supervising the services.

The role of the electorate is to act as a vetoing body. They are asked to support or reject a single proposal. It is not easy to draw welfare inferences from a community vote, since many cast their vote based on what they believe will be the reaction of the city officials. There will be some who prefer an alternative to the proposal on the ballot, and some of these will vote negatively hoping that the preferred alternative will be put on the ballot, while others will be dubious about the response of the officials and thereby support the measure. We will ignore the element of strategy of voting and assume that an individual casts his vote on a simple private benefits-costs calculation. If this evaluation of his private benefits exceeds his evaluation of his private costs, he will vote in support of the investment. Let us go on to how the individual can treat these benefits and costs.

The financial costs of an investment are relatively well known. Typically, the proposing government will publish the schedule of operating, maintenance, interest, and debt payments. Often these figures for the agency will be transformed into costs for a typical resident. The computation is relatively simple. The bulk of government receipts will come from the property tax. Given the real property of a typical owner-occupant resident and the payment schedule of the project, the

annual payment of an average resident can be announced and, of course, if only the expected tax rate is announced, any resident can estimate his own financial costs.

These financial payments do not provide information about the incidence of costs, but the residents generally interpret the payments as costs. In fact, some of the costs may be shifted, while others may be fully borne by present property owners.

The benefits are much less apparent. There are few investment projects which are used by all residents. Clearly, if we were debating the establishment of a fire-protection system, everyone would receive some direct benefits, but the usual investment is an increment to an existing system. Even though the increment may be large, it will only expand part of the system; e.g., a new fire station will be established, some ancient equipment will be replaced, etc. If possible, the proponents will try to distribute direct benefits as widely as possible, but usually this is not feasible. (Certainly, the childless couple will not directly benefit even if every school is given a swimming pool.) The absence of direct benefits does not mean that the benefits are non-existent. A freeway may benefit almost all street users, but the incidence of benefits will be even less well known than that of costs. Whatever the actual or fancied distribution of benefits and costs, each potential voter will make his own estimate and, let us assume, he will support the measure if in his judgment his benefits are greater than his costs.<sup>1</sup>

Perceived costs, we are saying, are annual charges based on the value of his real property. The value of his property is the sum of the present value of future incomes to be earned with the property (imputed incomes if the property is owner-used). The stream of individual payments to the government becomes a reduction in the present value of the property. The stream of benefits of the investment can be transformed to a present value to be compared with the present value of tax payments, but there are asymmetries in concept, in distribution, and in measurement between the benefits and costs estimates.

An important difference between benefits and costs is the perception of these two series. The financial costs, whether real or transfer, are assessed against each property owner; the benefits are relatively hidden. Few public services conform to the classic characteristic of public goods. A large part of public investment is an integral part of site development of specific parcels. Some investment may directly serve specific groups, but elsewhere it relieves congestion and therefore leads

<sup>1</sup>This basis for a vote is not obviously optimal for the individual. It means that an individual will support a proposal even though there is an alternative, feasible proposal which he believes preferable. A better basis for his vote would be what he believes will be the consequence of a revoting sequence if this measure is defeated, but this strategic logic would take us too far afield.

to quality improvements to others. A new neighborhood park may benefit all—it will reduce use densities in all parks—but few of the non-users of the new park will recognize their benefits. Costs are announced against all residents and benefits are clear for a few. In a longer run some costs will be shifted and benefits will become more dispersed, but few voters can or will assess long-run effects. The weight of the more compelling information, who will pay and who will use, will bias the choice towards a negative vote.<sup>2</sup> Therefore the perception of benefits and costs by individuals will inhibit support via referenda.

The costs are assessed proportionately. The benefits are not. For a vote to carry, it is only sufficient that one-half or two-thirds (or whatever fraction is required) of the residents judge their benefits greater than their costs. Two-thirds of the residents may judge that their benefits are greater than costs but, in the aggregate for the community, benefits may be below the costs. Due to the phenomena of a single vote which is not weighted by degree of gain or injury, it has been argued that there will be excessive public investment where a referendum is the technique of public choice. Since a major point of this paper is that the inclusion of referenda in local investment reduces the likelihood of locally sponsored investment, and therefore leads to the growth of investment by higher level governments, let us look more closely into this argument.

The negative and positive judgments of individual members of the community may be due to the locational incidence of the benefits (a road that serves only part of the community) or it may be due to difference in preferences (not all of us feel as strongly about the quality of education, a handsome civic center, etc.). Whether differences in evaluation are due to tastes or distribution of public goods will affect the analysis.

The case we want to explore is the referendum for a public good, let us say a civic center, which is carried by a majority vote but where the sum of felt losses (benefits minus costs) by the negative voters is greater than the sum of felt gains of the more numerous supporters. The first problem is the definition of "gain."

The benefit to the individual resident is usually defined as the stream of income he would surrender in order to have the product of the public investment. Since these benefits are provided freely to residents, he would add them to the value of his residence in the community. But the market value of the property does not go up by the full amount. The market value will rise because of the increase in demand for residence in the community due to the services of the public investment—

<sup>2</sup>How many workers would vote to support research on automation though in the long run their incomes thereby may be increased?



the civic center. Assuming the migrants have tastes similar to the residents, the increase that the migrants are prepared to pay is equal to the average that the residents are prepared to pay. Therefore, the increase in market value of the individual's residence is the average value given to the increased public service by all of the residents. The individual realizes a consumer surplus due to the difference between his evaluation and the market price.

In the specific case of the civic center where the majority voted "yes" but the felt losses were greater than the felt gains, the increase in demand for residences due to the public service would be less than the decrease in demand due to taxes, so that the aggregate of property values would decline. An interesting phenomena has occurred. The winners of the vote have suffered a loss in property values; so have the losers. The property losses are equally distributed around the community. The sum of property losses of the gainers is greater than the sum of the consumers surpluses received by the gainers. On selfish grounds, the men who would support the investment on the basis of a comparison of their evaluation of direct benefits to themselves and direct costs to themselves, would oppose the measure once the effects on property values would become apparent. These conclusions are based on an assumed constancy of distribution of tastes.<sup>3</sup> The change in property value of the supporters reflects the disbenefits of the opponents of the measure.

If the felt losses and gains arose because of specific locational effects, the above mechanism would not operate. Within the community, the handicapped areas will lose and the advantaged one will gain in value; the bond issue would be passed. But, as in the preceding case, the asymmetry between the perception of costs and benefits will more likely lead to a negative vote. The benefits are often indirect and poorly understood while the tax payments are immediate and apparent.

Let us turn our attention to another aspect of the individual evaluation which reduces local support via referenda. The referendum requirements of local investment overtly cast the decision into one of private calculations. For individuals, there are a flow of payments and benefits which they evaluate at the discount rate used in their private calculations. The costs of the investment are presented as a series of tax payments which are a function of the local government interest rate—typically varying between 3 and 4 percent. The discount factor of the electorate is at least twice that of the government. Certainly, if the

<sup>3</sup> A possible long-run consequence of the passage of the bond issue is that the losers would leave the community and their replacements would not be randomly chosen from the outside population, but they more likely would be persons who would have voted for the bond issue. Therefore, in the long run as the taste composition changes, there will be increases in property values.

individuals had found an investment advantageous to finance on their own account, they would have even more welcomed the government. But just as the present value of the costs is reduced, so is the present value of the benefits. The individuals will use their higher discount rate to evaluate the stream of benefits in making their voting decision. The consequence of this is that the constraint of popular referendum is likely to cause the project to be of a smaller scale than what would have been developed had it been left to the good graces of public officialdom.

Where local public investment is produced under constant costs, few problems develop. It tends to lag slightly behind shifts in demand. Crises develop, scandals emerge, and the city energizes itself. The more difficult planning questions develop around decreasing costs cases, where there are heavy set-up costs associated with investment and the incentives to expand the scale are great, such as in transportation and resources development systems. The planners, engineers and administrators, are very conscious of decreasing returns. Their image of full development is matched by the chamber of commerce view of dense settlement with all of the capital gains associated with development. But both views of public investment are views of distant horizons. The benefits are far removed in time. The temptations of optimistic expectations are countered by the needs of a local referendum and legislative constraints restricting bond issues to a percent of assessed valuation. This capital rationing feature and the more cautious behavior of residents dampen the plans of local governments and give rise to charges of myopia and underdevelopment.

One path out of the bondage imposed by democratic referenda and institutional constraints to inhibit "premature" development (a recognition that referenda are not a sufficient constraint on the optimism of community leaders) is the transfer of function to higher governments. Higher level governments are less subject to the same inhibitions which face local governments. First of all, we predicated that each individual calculated his benefit; but this is highly fanciful. Benefits are difficult to estimate and highly uncertain, whereas in the local governments assessment of costs are too apparent. The higher the level of government, the less obvious becomes the cost assessment and therefore the cost calculation for the individual. Not only do individual costs become lost, but plans become part of the daily operations of an engineering office which is more sophisticated about the description of benefits. We are all familiar with logrolling and other devices by which higher level governments vote support for local projects—efficient and inefficient. I want to comment on only one feature which is of special interest.

The higher the level of government, the lower the discount rate used

in project analysis. In many cases, the rate is lowered as a response to some of the above factors which accounts for a greater receptivity towards a project by a central government. The discount rate used by individuals in private calculation contains a large uncertainty factor. Typically, the greater the uncertainty, the greater will be the risk factor and the higher the discount rate used by individuals and the governments which are closest to them. But governments are not confronted by the stringent consequences of bankruptcy. Unreliability or income disappointment in a specific project need not have the same calamitous consequences that it has for individuals. In fact, one might assert that central governments tend to have a gambler's preference; i.e., instead of downgrading alternatives with higher income prospects and greater probabilities of loss, they may be willing to pay a premium. The investments most typified by uncertainty and limited by risk are the developmental projects which involve long planning horizons and usually heavy capital investment with great excess capacity in early years (water supply, transportation networks). From the social viewpoint, these projects should be evaluated at their expected values without special handicapping for risks.

Many other factors of government structure are important in explaining investment. Of great significance are the territorial limitations of local governments which inhibit them from extending the scale of projects, despite possible decreasing costs, and lead them to ignore benefits or costs external to their area. Solutions are not always optimal. Adjustments involve competitive behavior among governments of similar functions and among the different levels. Investment criteria and explanatory hypotheses of behavior will be a function of the structure of governments and their internal organization. The expansion of the concept of government will prove highly useful.

## DIVERGENCES BETWEEN INDIVIDUAL AND TOTAL COSTS WITHIN GOVERNMENT\*

By ROLAND N. MCKEAN

*The RAND Corporation*

It is a bit surprising to realize what different attitudes we have toward the private and public portions of the economy. This contrast is particularly marked with respect to divergences between individual and total costs or gains. By such divergences I mean differences between costs and rewards as perceived by decision-makers and total costs and rewards produced by their actions. These differences are often called external economies and diseconomies or, more briefly, spillover effects. They are impacts on others that are not taken into account by managers or individuals who take action.

Let me review our attitudes toward the two sectors in this respect. In the private sector we have assumed, first of all, that individuals are utility-maximizers. We have not claimed any precise knowledge of individuals' utility functions but have concluded that utility must be a function of many desired items, that there are trade-off or substitution possibilities among these items, and that, if one becomes more expensive relative to others, less of that item will be demanded. On this foundation, we have constructed most of economic theory, including many testable hypotheses.<sup>1</sup>

In this connection, we have believed that a producer recognizes costs he has to pay but is unlikely to recognize damages or resources used up for which he does not have to compensate anyone. He would have to sacrifice too many other things to make unnecessary compensations. For similar reasons we have assumed that a producer recognizes benefits for which customers compensate him but is unlikely to count other benefits that he does or could produce. Again it would be too costly in terms of other objectives. Indeed in a moderately competitive industry the sacrifice entailed by a highly altruistic attitude would be bankruptcy. We have not meant, of course, that people always pursue self-interest narrowly defined. Rather we have meant that the higher the cost of pursuing one objective, the less of it one will try to achieve. Thus the more personal gain one sacrifices to be altruistic, the less

\* I am greatly indebted to the Social Science Research Council for an award that made possible a larger study of governmental processes, from which the present paper stems.

<sup>1</sup> The clearest discussion of these points that I know of is in an unpublished economics textbook by Armen A. Alchian. See also Armen A. Alchian and Reuben A. Kessel, "Competition, Monopoly, and the Pursuit of Money," in *Aspects of Labor Economics* (N.B.E.R., Princeton Univ. Press, 1962), pp. 157-83, and Gary S. Becker, *The Economics of Discrimination* (Univ. of Chicago Press, 1957).

altruism he will pursue.<sup>1</sup> And the greater the divergence between individual and total costs and gains, the less likely one is to be guided by total costs and gains.

Second, with respect to the private sector, we have written and talked a great deal about ways to define property rights better or about other ways to intervene so as to induce people to take total costs and gains into account. In this writing and talking, a variety of issues has been explored. It has generally been accepted that it is not desirable for people to take pecuniary spillover effects into account. And some economists—especially Ronald Coase—have stressed that the cost of government intervention should itself be recognized—that divergences between individual and total costs in the private sector do not call for government intervention unless it would actually do more good than harm.<sup>2</sup> In any event, we have given quite a bit of attention to external economies and diseconomies in the private sector and what to do about them.

Third, it has not been uncommon to conclude that activities should be transferred to the public sector if serious divergences cannot be eliminated. As cities developed, for instance, it became obvious that private decisions had important spillover effects on other persons. Many concluded that the only thing to do was to have governments plan cities or, as cities grew older, plan their renewal. Flood control and education are also examples of activities that had significant external effects—divergences between individual and total costs and gains—and were placed to a considerable extent in the public sector. I am not suggesting that this was obviously a wrong solution—I am merely saying that we do often decide to turn to government in such instances. And even where lesser spillover effects occur, it is sometimes implied that, because the market system has “failed,” the activity should be conducted by government.

In the public sector, however, our attitude has usually differed from the above. First, many of us, in our work, have tended to assume, unconsciously for the most part, that public officials are public-interest maximizers. In other words, we have assumed that government personnel pursue one type of gain and avoid one type of cost—those felt by the general public—and ignore other variables that would normally be in utility functions. To be sure, we have recognized the existence of outright graft and corruption, attributing this to government personnel of another extreme variety—selfish-interest maximizers whose utility functions include a number of evil aims and little else. These evil men seem to be regarded as the exceptions that prove the general rule of

<sup>1</sup> Ronald H. Coase, “The Problem of Social Cost,” *J. of Law and Econ.*, Oct., 1960, pp. 1-44.

public-interest maximizing. For some reason we have tended to neglect the intermediate possibilities: government officials who are ordinary utility-maximizers, with many items in their utility functions, with substitution possibilities existing among these items, and with negatively-sloped demand curves for each of these items. (Or, alternatively, government officials who are assumed for analytical purposes to have a smaller number of specific, though not wicked, aspirations.)

Second, given the above attitude, economists have naturally enough talked little about manipulating the costs of objectives sought by public officials so as to affect their behavior. We have talked little about rigging individual costs and rewards in government to make them more nearly consistent with total costs and rewards. Students of public administration and political science have searched for improvements in governmental organization, of course, implicitly seeking ways to bring individual criteria closer into line with higher-level criteria. Economists, however, have not given much attention to bargaining processes in government, the resulting cost-reward structures confronting officials, or ways to influence those cost-reward structures.<sup>3</sup>

Third, and again it is natural enough in view of the preceding remarks, we have rarely concluded that, because of gross divergences between individual and total costs or gains within government, certain activities should be shifted to the private sector. Again, I am not suggesting that our policies have obviously been wrong. It simply strikes me as being a curious asymmetry in our thinking.

Or does each government official, in choosing his actions, weigh total gains to everyone against total costs to all? Is he unwilling to trade part of this objective for some degree of achievement along other lines? It may be true that public servants aim to promote the general interest more consistently than most other samples of the population. There is a selection process in the flow of personnel into any occupation, and

<sup>3</sup>In recent years, though, there has been growing interest in economics as well as other social sciences in studying organizational behavior from vantage points of this sort. Some of the more directly relevant efforts are Armen Alchian's work on the implications for behavior of various kinds of property rights; E. C. Banfield, *Political Influence* (Free Press, 1961); James Buchanan and Gordon Tullock, *The Calculus of Consent* (Univ. of Michigan Press, 1962); R. A. Dahl and C. E. Lindblom, *Politics, Economics, and Welfare* (Harper & Brothers, 1953); Anthony Downs, *An Economic Theory of Democracy* (Harper & Brothers, 1957); George C. Homans, *Social Behavior: Its Elementary Forms* (Harcourt Brace, 1961); C. E. Lindblom, *Bargaining: The Hidden Hand in Government* (The RAND Corporation, RM-1434, 1955); Mancur Olson, Jr., "A General Theory of Public Goods" (Mar., 1963, unpublished manuscript); William Riker, *Theory of Political Coalitions* (Yale Univ. Press, 1962); Jerome Rothenberg's work in process on models of government behavior; Herbert A. Simon, *Administrative Behavior* (2d ed., Macmillan Co., 1961); J. G. March and H. A. Simon, *Organizations* (Wiley & Sons, 1958); and Gordon Tullock, "A General Theory of Politics" (unpublished manuscript). Also, this way of looking at organizational behavior, while often neglected, goes back a long way. For a review of the interest of one famous political economist in this subject, see Nathan Rosenberg, "Some Institutional Aspects of *The Wealth of Nations*," *J.P.E.*, Dec., 1960, pp. 557-70.

those who choose public life may well be less selfish than most others. Nonetheless, it is unlikely that there is a sharp distinction between these two populations, especially since there are numerous transfers back and forth. Government officials too are surely utility-maximizers. They may attach less weight than others do to personal costs and gains and give greater consideration to the costs and benefits bestowed on strangers. But still, the greater the cost of achieving one aim, the less of it they will try to achieve; the greater the divergence between private and total costs, the less likely they are to reach decisions in the light of total costs. Moreover, to stay in office or to survive in the bureaucratic struggle, one cannot ceaselessly strive for the public interest. As in the competitive model, the cost is bankruptcy—this time career bankruptcy. The best one can hope for is that on balance he is “doing good” while compromising on many individual issues. A public official may not ask, “What can GNP do for me today?” but neither does he ask each morning, “What can I do for GNP?”

Even if civil servants and politicians do not constantly look at total costs and gains, however, are there serious divergences in government between private and total costs or gains? After all, there is a mechanism in any organization that forces each member to take into account many effects that might otherwise be externalities. This mechanism is the bargaining process, and it is crucial in making costs and gains felt by decision-makers in the public sector. It is similar in certain respects to the price mechanism, which is crucial in making costs and gains felt by decision-makers in the private sector. When a business firm takes action, it has to bargain with and compensate numerous persons who supply buildings, labor services, and other inputs. That is, if the firm's action uses up or damages property, the firm has to buy the consent of the owners. Wherever the firm's action has beneficial effects, the management tries to charge the beneficiaries. The greater the extent to which all these compensations are made, the less the extent to which the firm's costs and gains will diverge from total costs and gains. In government, similarly, if one official's action will use up someone's property or damage their interests, the official will probably find a cost associated with the action. He may endure embarrassing or expensive enmities, or suffer costly retaliations. One way or another he will have to bargain and pay a price, the size of which will depend on bargaining strengths and circumstances. If an official's action benefits a colleague or group, he will be able to bargain, tacitly if not explicitly, for some kind of *quid pro quo*. Thus the bargaining process can work in the direction of making individual costs and gains more nearly reflect total costs and gains.

In this grossly imperfect competition, however, the process does

not work with much precision. Individual consumers have no incentive to put much effort into bargaining, for example, while advocates of tariffs, silver subsidies, or price-support programs have big stakes and develop powerful bargaining strengths. The process may lead to desirable things part of the time, for logrolling may help minorities protect their rights or achieve good things for which they are willing to pay a high price. In secret ballots, which could prevent any trading, majorities might well make laws to suppress individual rights for various minority groups. Perhaps many uneconomic or inequitable policies are simply part of the cost of protecting individual rights in a majority-rule society. But the main point is clear: this "price mechanism" in government performs most imperfectly. Often groups that in the aggregate are affected greatly, can bargain only weakly, and the "price-tags" that become attached to various actions are far from the "right" mark. Thus costs and gains as felt by officials can indeed diverge, and perhaps seriously, from total costs and gains.

It is fairly easy to think of examples. Consider the lease of rights to drill for oil in the deep waters offshore from Los Angeles. The City Council awarded such a lease several years ago. From the standpoints of at least a majority of the Council, the gains from this action outweighed the costs. When Mayor Yorty was elected, however, he felt differently. In fact he had the City Planning Commission zone most of these deep waters "for residential use only" to prevent any drilling; and then apparently considered awarding offshore leases for other ocean areas. In such situations it is hard to believe that everyone is looking at costs and rewards from the same overall viewpoint. It reminds me of the typesetter's Freudian slip when a New York paper stated in a different situation: "This will not hurt the mayor, some feel, in that he can say he was always on the side of the angles."

More seriously, though, in connection with almost any government decision, it is instructive to try putting yourself in various officials' places—Senator A, Governor B, Secretary of Department X, Chief of Branch Y—and imagining the gains and costs of alternative actions as you would perceive them. The magnitudes cannot be measured, but judgment suggests that serious divergences are pervasive, that checks and balances are crucial in preventing the outcomes from being disastrous, and that it may be possible to devise better bargaining arrangements yielding significantly improved cost-reward structures.

Perhaps the most important divergences between individual and total costs stem from the fact that government officials are spending other people's money. Almost no participant in the budget-formulation process is guided by a cost structure that is in line with total costs. The deck appears to be stacked in favor of gradual budget growth. First,



there are groups of firms and individuals who find it worth while to press for favors. (One of the failures of marginal productivity theory is to consider the marginal productivity of effort devoted to obtaining favors from government.) Second, government personnel find these and other expansions attractive. For example, if the head of a bureau or department gets an increased budget, he reaps significant rewards: he can do a better job, or satisfy pressure groups, have greater influence, increase his chances for advancement, and so on. The costs include mainly effort devoted to appeasing rival department heads (usually by making his strategy consistent with the growth of other departments too), and efforts devoted to getting congressional support. If anything akin to the real resource cost enters into the calculations, it must do so by way of constraints from above.

When we turn to control by top levels of the executive branch, however, we find that here, too, spending other people's money pays. A government can win support by spending in strategic places and spreading the cost thinly over a large group of taxpayers. Tacit coalitions arise, and the restraint once exercised by Treasury Control in Great Britain and by the Bureau of the Budget here is gradually eroded.<sup>4</sup> But what about the top legislative body? Do members of Congress feel the real resource costs and transmit to government a corresponding constraint? No, as long as they do not go wild, they too can get more points by judicious spending than by voting against tax increases. Imagine that you are a senator, and compare the points you would score if, apparently singlehanded, you promoted a home state river project with the points you would rack up if you, along with several hundred other representatives and senators, reduced each voter's tax bill by \$10.

So much for this fragment of cost-reward structures within government. What about the taxpayer? Doesn't he feel the real resource costs and bargain for restraint? Let us look at the costs and rewards, from his standpoint, of pressing for tax reduction or opposing a tax increase. The gains from success in such an endeavor might amount to \$100, but the probability of an individual influencing the outcome is infinitesimal. The expected gains from an individual's effort to oppose a tax increase, therefore, are virtually zero. In these circumstances, how much effort will he devote to bargaining?

Thus it may be that no one feels a cost that adequately reflects the

<sup>4</sup> See W. Drees, Jr., *On the Level of Government Expenditure in the Netherlands after the War* (H. E. Stenfort Kroese, N.V., Leiden, 1955), pp. 61-71; Alan T. Peacock and Jack Wiseman, *Growth of Public Expenditure in the United Kingdom* (Report of the N.B.E.R., Princeton Univ. Press, 1961); and Alan T. Peacock, "Economic Analysis and Government Expenditure Control," in Alan T. Peacock and D. J. Robertson (eds.), *Public Expenditure Appraisal and Control* (Oliver & Boyd, Edinburgh, 1963).

real cost of budget increases. It is hardly surprising, therefore, to find that you as taxpayers are helping to put up a million dollars to rebuild the town of Wink, Texas, and half a million dollars to build a stadium in Bridgeport, Connecticut. Some people regard the latter example as being symbolic because Bridgeport was the original home of P. T. Barnum. But even if you believe such federal subsidies are wrong, no one is necessarily being a sucker, and no one is necessarily behaving in a reprehensible fashion. Gradual expansion of central government is probably where utility-maximization leads with the divergences between individual and total costs that exist under present institutional arrangements.

What can and should be done to reduce such divergences? Many devices may be worth considering: e.g., having numerous congressmen-at-large, elected by the whole nation; agreeing, because of our long-run interests, to have paid oppositions; or agreeing to have a "taxpayers' union," with closed shop and check-off system. Perhaps most of all at this point, though, we need to acquire a better understanding of the costs and rewards that organizational personnel find attached to alternative choices. In our familiar competitive model with its cost-reward structures, utility-maximization leads to fairly good outcomes. When you put governmental units (or an international community of governments!) into the model, however, where does utility-maximization lead? And where would it lead under modified arrangements affecting the various divergences between individual and total costs and gains?

## DISCUSSION

MANCUR OLSON, JR.: One of the characteristics that the interesting papers presented at this session have in common is that they appear at first to fall outside the boundaries of economics. The papers by Drs. Buchanan and Margolis deal primarily with voting, while Dr. McKean's paper is concerned mainly with politicians and bureaucrats. The subjects of these papers therefore fall squarely within the domain of political science. Yet we economists can see in these papers the signs and skills of our craft. And it is probably the style of thought more than the substance of the subject that distinguishes economics from the other social studies. Economists, in any event, should use their skills to study whatever problems they have a comparative advantage in studying. Thus these authors deserve our thanks for bringing the economist's intellectual machinery to bear on these important and long-neglected problems.

Another characteristic that these papers have in common is a concern with the way political processes determine the amount of resources allocated to the production of public goods. Buchanan and McKean suggest that the American political and fiscal system systematically tends to allocate excessive resources to the satisfaction of social wants, while Margolis suggests that local referendums, at least, can lead to less than optimal levels of public investment.

The place to begin any discussion of the merits of these different views is with Buchanan's model. In its beginning form, this model assumes both identical tastes and identical incomes, and in such circumstances an optimal amount of the public good is allegedly purchased, whatever the type of tax, so long as there is horizontal equity. This argument is not strictly correct, for to the extent that a tax distorts the allocation of resources it will have an "excess burden" that will lead to a smaller purchase of the public good. Extra purchases of the public good will entail not only a direct resource cost, but also any extra burden imposed by the tax's tendency to distort the allocation of resources. To the extent that a given tax leads to a less efficient allocation of resources than a lump-sum tax, it will lower income, and therefore normally lead to a different level of purchase of the public good. This minor point is not, however, essential to Buchanan's principal argument.

It is nonetheless important to point out that the possible tendency toward excessive purchases of public goods in Buchanan's final model may well be counteracted by opposing factors. First, demonstration effects are surely strongest in the private sector, and they tend to lead toward an excessive purchase of private goods. Second, billions of dollars are spent advertising private goods, while public goods are normally not similarly advertised.

A third, and most important, factor that can make the provision of public goods inefficiently small is evident in the McKean and Margolis papers. McKean tells us that a congressman is more apt to vote for a home state project than for a tax cut. This is presumably true, but while a tax cut will

have some appeal to congressmen from all fifty states, the home state project as it stands will interest only those from the one state. There will need to be logrolling before the home state projects, however worthy, will get majority support. This situation is characteristic of local government as well, as Margolis points out. He emphasizes that most of the benefits of typical local government investment are concentrated on some subset of the community's population. It is probably inherently the case that the benefits of a single government project are usually less widely dispersed than the taxes that finance it. It follows that many proposed government projects are such that, while the gains from undertaking them would exceed the losses, the distribution of the gains and costs is such that these projects will not by themselves command a majority vote. Thus desirable additions to the stock of social goods can often be made only if there is logrolling.

But logrolling is a costly, cumbersome, complex, and uncertain procedure. There is no assurance that the logs will be rolled, even when there are general gains to be gotten from rolling them. Logrolling may be relatively easy among a relatively small group of senior senators paid to perform that function. But it may be very difficult in many other cases, especially when popular elections are involved. A group that would gain from a given government project may find that even the lobbying organization needed to bargain with other interests is, to this group, a public good. As a result there may be no incentive for individuals voluntarily to support the institutions that would make logrolling feasible. So the logrolling needed to support many Pareto-optimal public investments may not take place.

Even when there is no need for logrolling, it may be difficult to get agreement on the purchase of a public good. An individual buying private goods can buy as much as he wishes without consulting or coordinating with anyone. But when a public good is to be purchased, agreement must first be reached on how the costs are to be shared. It is often a good bargaining tactic to withhold support for a desired public good in an attempt to reduce the share of the cost of the public project that one has to pay. A Pareto-optimal public project may therefore be undertaken only after much bargaining and delay, and then on a less than optimal scale.

There are, to be sure, still other and so far neglected factors affecting the supply of public goods in one way or another. Thus one hopes that other economists will use their theoretical skills to study this problem, the way the authors of these three stimulating papers have done.

NATHAN ROSENBERG: The history of the economist's treatment of political behavior and policy making presents a curious and not entirely edifying spectacle. The eighteenth-century origins of our discipline are steeped in a conception of human behavior which was uncompromisingly self-interested and egoistic. The keynote was perhaps set by Mandeville, who assured his readers that "there is nothing so universally sincere upon Earth, as the Love which all Creatures, that are capable of any, bear to themselves." As a counterpart to this view, the possibility of disinterested, to say nothing of altruistic, behavior on the part of all public officials was treated with at

least skepticism and more often contempt. Adam Smith made withering references to "... that insidious and crafty animal, vulgarly called a statesman or politician," and David Hume accepted it "... as a maxim, that, in contriving any system of government, and fixing the several checks and controls of the constitution, every man ought to be supposed a *knave*, and to have no other end, in all his actions, than private interest." Yet somehow or other economists allowed themselves to be saddled with a Benthamite legislator whose own preference function played no role in political decision making and whose adroit use of the felicific calculus enabled him to calculate that wondrous ambiguity, "the greatest happiness for the greatest number." Perhaps even more astonishing than his ability to make this calculation was the assumption that he would in fact legislate accordingly.

The contributors to this session have all participated in the recent reappraisal by economists of political processes, and in the attempt to analyze political decision making in a manner at least more consistent with traditional economic theorizing. Such a reappraisal is long overdue and is, therefore, to be welcomed. Economists have for too long confined their activities in this area to a Pigovian search for discrepancies between marginal private and marginal social costs and benefits, assuming rather casually that, once such discrepancies in private markets have been uncovered, all that remained was to correct them via some appropriate tax or subsidy scheme. McKean has, quite correctly, argued that the Pigovian analysis is also applicable to decision making in the public sector, and that the discovery of non-Pareto optimal conditions in the private sector does not suffice as a case for public expenditures. In this respect his work is in substantial harmony with the earlier work of Coase and Buchanan, and his plea for an exploration of the manner in which the institutional arrangements of the government sector create divergences between private and social costs is one which I would heartily endorse.

On the other hand, Samuelson's recent work in the pure theory of public expenditures and the discussion centering around it have made it apparent that we can no longer confine our fishing even to the somewhat muddied waters of the Pigovian stream but must explore the almost impenetrable depths inhabited by those numerous fish possessing some of the markings of a public good. To be sure, pure public goods are extremely rare, aside from that awesome whale of national security and those elusive minnows of new knowledge. But as soon as we consider the whole range of decreasing cost phenomena, of which Samuelson's pure public good is simply an extreme case (i.e., the marginal cost of providing some benefit to an additional user is literally zero), our fishing expedition becomes one of a frenzied rush to haul in the nets. For so long as goods have some element of publicness attached to them, even an otherwise perfect market fails to achieve Pareto-optimal resource allocation.

Furthermore, when we relax the independence postulate of welfare economics, as I think we must, the limitations of our traditional approach become even more apparent. My welfare is affected not only by the much-

discussed disutility of envy but also by the poverty of others, whether this takes the form of social disturbances generated by urban slums or the potential dangers of living in a society where emotionally-disturbed people are too poor to afford the costs of psychiatric care.

It appears, then, that not one but two Pandora's Boxes have been opened, and we are confronted with highly disconcerting real world choices between imperfect markets and imperfect governments. It seems to me that a major task of the economist is to help clarify both the analytical and organizational issues involved in making such public versus private choices, and also to enlarge public awareness of the range of alternatives available and their implications. On this score I am somewhat unhappy over the sharpness of the dichotomy between the public and private alternatives present in McKean's paper. These organizational polar extremes are exactly analogous to the purely public versus purely private goods of our analytical models. Just as the difficult theoretical problems are involved in that large gray in-between area of goods which bear varying degrees of publicness, so the organizational alternatives involve varying degrees and kinds of public participation. Not only do we have public production of the TVA variety but we also have public regulation and control of utilities, public subsidy of private enterprise, public highways financed on a user cost basis, public and private educational institutions living (almost) happily together, public support of medical research which is simultaneously financed partly from private sources, and public sponsorship of specialized activities contracted to private organizations—the Atomic Energy Commission and even the RAND Corporation. Needless to add, even this list is suggestive rather than exhaustive.

This point poses a closely related issue. In clarifying the alternatives available to society in achieving an efficient utilization of resources, the complex of institutions in both the public and private sectors requires searching re-examination. McKean has correctly pointed to the dysfunctional aspects of political decision making which result from the peculiar structure of the cost-benefit system as it confronts a subjectively rational decision-maker in government. But self-interested behavior always takes place within some institutional matrix which determines the selection of variables to which the individual responds as well as the nature, direction, and size of the response. And just as it is an error to contrast market allocations with the decisions of Benthamite legislators, so is it equally misleading to contrast government allocation mechanisms with that of atomistic markets where such markets no longer exist. The "organizational revolution" of the twentieth century has had its impact not only on government but on the private sector as well, where decision making frequently resides within the executive structure of large corporate bureaucracies which bear many of the earmarks of government bureaucracies. I do not want to overstate these similarities. But it seems apparent that the incentive system confronting a senior executive in General Electric is not the same as that confronting a small family-owned retailing operation. The signaling devices, inducements, pressures, and constraints to which he responds are

not immediately market determined but are established by his role and opportunities for promotion as established by a large corporate hierarchy. It is at least possible that his private cost-benefit calculations also contain dysfunctional elements not present in a small business context where decisions are more directly mediated by market forces. Examination of these problems involves an approach fundamentally different from traditional welfare economics and more in accordance with the tentative steps recently taken toward the development of a theory of organizations.

My final remarks take me beyond the realm of economics narrowly defined, as it seems to me is inevitable in any discussion of public expenditures. First of all, I think we must abandon the approach according to which the consumer is pictured as possessing a pre-established, autonomous, preference structure, thereby reducing the problem to the comparative merits of public versus private expenditures in catering efficiently to these wants. Indeed, it is difficult to see what political leadership, political debate, and public discussion are all about if wants are already rigidly fixed. The public dialogue must be regarded, at least in part, as a device for informing individuals in a democratic society and thereby enabling them to formulate and evaluate alternatives which they did not previously perceive. Such discussion performs the educational function of enabling the public to examine possible policies which they had not previously considered either because they were badly informed or incapable of analyzing complex economic interrelationships.

Finally, the free market is a highly imperfect mechanism for registering consumer preferences for such things as stability, security, and the elimination of the private and social hardships associated with poverty. Nor is it an effective mechanism for eliminating either its own defects or personal or environmental defects which hamper the individual's capacity to respond to market forces. In this respect, many government expenditures represent an embodiment of the social goals and values of a democratic society as well as measures designed to improve rather than supplant the functioning of the market by raising the level of vocational skills, increasing mobility, and providing information. Insofar as there is a general consensus on the desirability of goals which cannot be achieved through the instrumentality of the market, discussions of the comparative efficiency of governments versus markets are not entirely in order. The critical question is then the selection of the most appropriate technique available to government and a careful scrutiny and weighing of the possible costs and consequences which such techniques may entail.

JEROME ROTHENBERG: The three papers have an important common focus: sources of suboptimality in public decision making. McKean addresses himself to the problem in general, concentrating on motivations of public officials and on the fact that *quid pro quo* financing is inappropriate. Buchanan concentrates on an aspect of the latter, that taxation is based on formulae other than individuated benefits. Margolis concentrates on biases in benefit-cost evaluations at different levels of government. Complementarity is present and is a welcome augur of significant illumination on

a subject becoming recognized as one of great importance. There is much I should like to discuss in these stimulating papers, but I am afraid I shall have to restrict myself almost entirely to Buchanan's.

In Margolis' paper, it is not clear how a divergence between private and government rates of discount imparts a negative bias against projects proposed in referenda, if both benefits and costs are similarly reduced thereby. Is the resort to higher levels of government spurred more by lessened tax awareness than by the hope to shift some costs to outsider nonbeneficiaries? On the other hand, isn't a deterrent to such resort that proportionate opposition will be greater and support less, since benefits are still locally concentrated, but costs are to be spread to many more beneficiaries?

Buchanan's paper presents a very interesting analytic apparatus, and it employs this apparatus to draw two conclusions about a decrease in tax progressivity: (1) it will tend to decrease the size of the public sector; (2) it will tend to improve the efficiency of resource allocation between public and private sectors. The first is a positive and the second a normative conclusion. Both can be subject to question.

As to the first, the main determinant of the size of the public sector in Buchanan's model is not progressivity as such, but the tax cost to the median voter. Outcomes seem to depend on the position rather than shape of the tax function; but progressivity is a function of both. A rotation of the tax function, with no shift, will also change progressivity. So one can easily reduce progressivity and evoke either increased, decreased, or unchanged public services, depending on the particular way that the reduction is accomplished. It is misleading for Buchanan to suggest that his model makes unambiguous inferences about the level of public services on the basis of changes in progressivity per se.

The problem is more serious with regard to the normative conclusion. The conclusion depends crucially on a certain degree of progressivity, represented in Figure 2 by  $t'_1$ , serving as a Pareto-optimal reference level. Thus it is argued that with a tax structure such as  $t_1$ —which is suboptimal—too many public services are produced and that, therefore, when progressivity decreases in such a way that the tax function shifts leftward toward  $t'_1$ , the outcome moves toward the optimum. The problem is that  $t'_1$  cannot in fact serve as a normative reference point; and it is difficult to discover a function which can. Thus movements toward or away from  $t'_1$  cannot be given a simple welfare interpretation.

Basic to the alleged role for  $t'_1$  is that it is derived from the optimal proportional tax  $t_1$  with the following properties: (a) it is linear; (b) it is symmetrical—it leaves the original median voter unaffected; (c) it leaves the majority decision unchanged; (d) it raises just the same revenue as does  $t_1$  for the chosen level of services; (e) these properties hold for any distribution of income. The reason for the linearity requirement is as follows. By switching from  $t_1$  to  $t'_1$  nonmedian voters are pushed into disequilibrium, since their tax price shown in  $t'_1$  diverges from their marginal evaluation shown in  $v_1$  by definition. But the resulting gains by the poor (measured by the divergence between the two curves) will just offset the losses by the rich; so that no bribes can be made to shift the situation elsewhere. This is



crucial to the alleged Pareto optimality of the  $t'_1$  situation. Curvature in  $t'_1$  would make for nonoffsetting gains or losses and thus make successful bribes possible.

The criticism is that neither linearity nor an undisturbed majority choice can be guaranteed. As to linearity, it can be shown that the existence of a linear  $t'_1$  depends on the distribution of income—a variable never introduced by Buchanan. For symmetrical distributions, where median income equals mean income, there is no problem. But symmetry is inappropriate for real world income distributions, where significant skewness exists. For asymmetrical distributions, linearity does not exist independently of particular distributions. That is, a linear  $t'_1$  can be found only for very special distributions.

[Proof. (I am indebted to Professor Meyer Dwass for verification of this.)  $Y$  is a random variable (income),  $M$  its mean,  $m$  its median.

$$(1) \quad t_1 = aY \qquad (2) \quad t'_1 = bY + c$$

We require:

$$(3) \quad am = bm + c \text{ intersection at median, and}$$

$$(4) \quad E(aY) = E(bY + c) \text{ same total tax revenue}$$

So from (4):

$$(5) \quad aM = bM + c$$

Solving (3) and (5) for  $b$ :

$$(6) \quad a(M - m) = b(M - m)$$

$$(7) \quad b = \frac{a(M - m)}{(M - m)}$$

So, if  $M = m$ , we have the singular case of trivial solutions. If  $M \neq m$ , we have only our original  $t_1$ : no other linear function is possible.]

Second, Buchanan's model does not guarantee that  $t'_1$  will keep the majority decision unchanged from  $t_1$ . The error lies in confusing the situation of the voter who had the median preference under  $t_1$  with the position of median preference. Under single-peaked preference majority rule, the majority choice depends on the latter, not the former—contrary to Buchanan's model. The situation of the erstwhile median preference voter may remain unchanged, yet if, as a result of unequal shifting preferences by others across the unchanged preference of this voter, this preference no longer is the median preference, the median will have shifted, and with it, the majority choice.

Just such unequal shifting is possible in the change from  $t_1$  to  $t'_1$ , since introducing progressiveness changes tax prices disproportionately to income. So the assumed elasticity properties of the valuation curves make likely that the demand by some voters above median income will drop below and the demand by some voters below median income will rise above

erstwhile median demand. Whether median demand changes depends here too on the income distribution. But only in very special cases (e.g., symmetry again) will the median remain unchanged.

If the median demand does change, by the way, another of the conditions on  $t'_1$  will be violated: taxes collected from  $t'_1$  will not equal the cost of providing the  $t_1$  level of public services. This leads to real stability problems. Assume, for example, the shift from  $t_1$  to  $t'_1$  induces a lower majority demand. Then, if we want to preserve Pareto optimality by keeping the new median voter in equilibrium, we must shift the tax function upward. But this will raise too much revenue. A linear rotation around the new median may not solve this, for reasons given above. So nonlinearity may be required; and this violates the Pareto-optimality requirement.

In sum, the model is incomplete for predicting majority choice; and the construction of an optimal tax situation as normative point of reference does not succeed. The true position and shape of such an optimal situation is not correctly specified. Normative conclusions based on such a construction are therefore unfounded.

# PROBLEMS IN TAXATION

## DIRECT SHIFTING OF THE CORPORATION INCOME TAX IN MANUFACTURING\*

By CHALLIS A. HALL, JR.  
*Yale University*

Traditional conceptions that corporation profits taxation does not alter marginal cost at the most profitable output in the short run and therefore does not change the relative inputs of labor and capital or the marginal contributions of these resources to production and revenue have been increasingly subjected to criticism in recent years. Some critics maintain that corporate enterprise regards the tax as an excise on output, and that the immediate impact of the tax is analogous to the direct effects of sales taxation in either raising output price corresponding to previous money payments to labor and capital at existing output or reducing receipts available for distribution to both labor and capital at given output market price. Other critics of the traditional view contend that the principal short-term impact of the tax may be to exert downward pressure on wage rates at any given output price and production level.

Resolution of these conflicting views—of considerable importance in understanding and evaluating the total consequences of the tax for public policy decisions—has not been achieved by previous attempts involving comparison of such variables as profit margins on sales, profit rates on invested capital, the profit share of income originating in the corporate sector, and so on.<sup>1</sup> Another attempt to investigate short-term shifting therefore appeared desirable, and a new approach which deals directly with wages and profits as productive contributions of labor and

\* This paper summarizes one phase of an inquiry into the corporation income tax and economic growth, sponsored by the National Bureau of Economic Research and assisted by a grant from the Rockefeller Brothers Fund to the National Bureau. However, the findings reported herein have not yet undergone the full critical review accorded the National Bureau's studies, including review by the Board of Directors, and are therefore tentative and provisional. I am greatly indebted to many individuals for advice and assistance on this research, especially to Robert Kilpatrick, now at Cornell University, for his laborious and dextrous work as research assistant on the project, and to Professor Robert M. Solow, of the Massachusetts Institute of Technology, but these individuals bear no responsibility for the findings. The research for this project was undertaken at the Brookings Institution during a leave of absence in 1961-62 from my duties at Yale University. The writer intends to introduce a more general approach in this research.

<sup>1</sup> The recent study by Marian Krzyzaniak and Richard A. Musgrave, *The Shifting of the Corporation Income Tax* (Johns Hopkins Press, 1963), was not published in time to permit comparisons with the results outlined in this paper.

capital appeared to offer promise. This paper outlines the methodology and results of this undertaking.

### *Method*

Orthodox economic analysis typically stresses the response of output to inputs of productive resources—especially physical productivity at the margin of use—as an important determinant of price, output, and the division of income between productive resources. The production function, which specifies the relationship between resource inputs and output, may in principle be derived for various types of labor and capital goods inputs with different degrees of productivity and for different types of output; but in analyzing broad sectors of industry, it is customary to combine the different outputs into a total production index, the different labor inputs into an aggregate, and various capital categories into one total.

Empirical demonstration of an aggregate relationship between output, labor input, and capital used is complicated by the continuing impact of technical change. However, Professor Robert M. Solow in a pioneering effort of a few years ago showed that technical change (broadly interpreted to cover improvements in labor quality as well as advances in technical knowledge) could be isolated and that a production function consistent with economic analysis could be derived from observed data for the whole private nonfarm economy and covering the years 1909–49 with a method based upon these hypotheses: (1) real earnings of labor and capital are determined by their marginal productivities and (2) given the state of technology, equal proportionate increases in both labor and capital used generate the same proportionate expansion of output.<sup>2</sup> In correcting output for technical change, Solow's original analysis made no *a priori* assumption about the character of technical advance but concluded that a special test indicated that technical change was neutral, in the sense that it raised the marginal physical productivities of labor and capital in the same proportion, at any given capital-labor ratio.

One attribute of Solow's analysis suggested that extension of its method could be fruitful in examining direct shifting of the corporate tax. Like most statistical analyses of productivity and absolute and relative factor shares in income, Solow's method had been used with the traditional view that profits taxation represents a charge against

<sup>2</sup> Robert M. Solow, "Technical Change and the Aggregate Production Function," *Rev. of Econ. and Statis.*, Aug., 1957, pp. 312–20. Some anomalous results were discovered to be computational errors in a critique of the method. Cf. Warren P. Hogan, "Technical Progress and Production Functions," *Rev. of Econ. and Statis.*, Nov., 1958, pp. 407–11, and Robert M. Solow, "Reply," *idem.*, pp. 411–13. Solow's technique of isolating technical change was applied to manufacturing by Benton F. Massell, in "Capital Formation and Technical Change in United States Manufacturing," *Rev. of Econ. and Statis.*, May, 1960, pp. 183–88. Massell, however, did not derive a production function.

capital: the output share attributable to capital was taken as observed profits including profits taxation. The method, in other words, inferred that the observed capital share represented the productivity of capital. However, if the corporate tax is, in fact, shifted in the short run, the observed capital share inclusive of tax exceeds the true productivity of capital by at least the amount of shifted tax. Reducing the observed capital share by an amount corresponding to shifted tax would lead to a different production relationship than that derived on the no-shift assumption. Would the production relationships derived under alternative shifting assumptions with the same set of basic data indicate a preferred tax-shift assumption? To answer this question, the method was applied to manufacturing, which is dominated by corporate activity and accounts for a substantial fraction of federal corporate tax receipts.

Viewed broadly, the method employed here to compare alternative shifting hypotheses consists in deriving a production function, corrected for technical change, for each shifting assumption, and then examining the internal consistency of these relationships as estimators of output and property income.

### *Derivation of Production Function*

Deriving a production relationship from time series data requires that the influence of technical change on output be removed. The procedures used in this process of adjusting actual output to obtain deflated output are outlined below in a step-by-step and discursive fashion. The steps followed here are: (1) measure the change in output per man-hour which occurs as a result of the cumulation of technical progress through time and (2) derive a relation between output per man-hour, deflated for technical change, and capital per man-hour. At each stage, the procedures are carried out for the traditional no-shift assumption and two alternative assumptions involving direct shifting, and the last step thus yields three different production relations.

Measuring the change in output per man-hour due to technical progress is accomplished by attributing some of the change in hourly output to capital deepening, if capital used per man-hour increases, and the remainder to technical progress. The attribution to capital is based on the assumptions of marginal productivity distribution and a linear, homogeneous production function, as specified earlier. For instance, if the share of property income in total output is 30 percent and capital inputs rise from one year to the next by 4 percent, labor input being unchanged, then 1.2 percent (.30 times .04) is approximately the relative increase in output assignable to capital deepening; if output actually rises say by 5 percent, 1.2 percent, or 24 percent of additional output, is attributable to capital deepening, and 3.8 percent, or 76 per-

cent of extra output, to technical change.<sup>3</sup> The formula for determining the approximate relative increase in output per man-hour due to technical change is:

$$(1) \quad z = \frac{\Delta q}{q} - s \cdot \frac{\Delta k}{k}$$

$q$  = output per man-hour.

$k$  = capital in use per man-hour.

$s$  = inferred marginal product of capital times capital per unit of output.

$z$  = approximate relative increase in output per man-hour due to technical change.

$\Delta$  = year to year difference.

The data used in the calculations of  $z$  and other variables cover the years 1919–59 and necessarily reflect various compromises and concessions. Output is defined to be gross real value added (income originating in manufacturing plus capital consumption allowances), but an index of output is employed in the absence of a real value-added series. Labor input is represented by total man-hours of employees. Capital in use is measured by the gross real capital stock (privately-owned plus government-owned-contractor-operated) corrected for the estimated time that the stock was not actually used in production. One definition of capital is comprehensive, including plant, equipment, and inventory. Calculations were also made on the basis of another real capital variant that includes only plant and equipment, since the stock in use corrections for fixed capital were different from those applied to inventory in the comprehensive concept. Capital's share in output  $s$  is the value in current dollars of gross profits observed (including profits taxes to the extent required by the shifting assumption and always including capital consumption allowances) divided by gross value added. Adjustments were made to exclude values attributable to foreign operations and to nonmanufacturing activities where necessary, but complete correction was not feasible.<sup>5</sup>

Capital's inferred contribution to output ( $s$  in equation (1)) depends upon the specific tax shifting hypothesis postulated, though in each case observed values are utilized for the productivity calculation. To examine the consequences of extreme assumptions, the inferred imputations

<sup>3</sup> Improvements in labor force, labor speedups, labor slowdowns, and advances in technical know-how are thus all picked up in this measure of technical change.

<sup>4</sup> Cf. Robert M. Solow, "Technical Change and the Aggregate Production Function," *op. cit.*, p. 313. Solow's formula has been modified in equation (1) to allow for alternative tax shifting hypotheses.

<sup>5</sup> Space restrictions necessarily preclude a detailed discussion of sources and estimating procedures in this paper.

to capital were calculated on the hypotheses of no-shift, full sales tax shift, and full wage shift. The consequences of these alternative assumptions for aggregate cost-price relationships and for the observed share of output attributable to capital productivity may be illustrated by an example.

Suppose observed values of output and its components in a given year are as follows:

Employee compensation (wages)	\$75 billion current prices
Gross profits (profits before-taxes plus capital consumption allowances)	35 billion current prices
Profits taxes	10 billion current prices
Current value of output	110 billion current prices
Implicit market price level	1.10 dollars
Real output	100 billion constant dollars

If the contribution of capital to output is postulated to be actual earnings on capital, adjusted for possible tax-induced distortions, evidently the traditional no-shift assumption requires that 31.8 units of real output be attributed to capital ( $35/110 \cdot 100 = 31.8$ ). For this case, the  $s$  in formula (1) is .318, or observed gross profits as a fraction of output value. If it is assumed that the tax is completely met from wages, and that property's contribution to output is its ex-tax amount, about 22.7 units of real output are attributable to capital ( $25/110 \cdot 100 = 22.7$ ). Alternatively, if business enterprise regards the profits tax as a sales levy on output, profitability of resource use is reckoned on the factor cost price level, which is market price level 1.10 less the tax per unit of output .10, or 1.00; and the value distributed to resources is output at factor cost ( $100 \cdot 1.00$ ), which happens to be equal to real output in the example. On this interpretation, the aggregate payment for use of capital is \$25 billion current prices, and since \$100 billion current prices at factor cost were produced, capital contributed 25 percent to real output ( $25/100 = .25$ ). Thus the share of output attributable to capital (i.e., the  $s$  in formula (1)) would be .318 for no direct shifting, .25 for full sales tax shift, and .227 for full wage shift.<sup>6</sup>

<sup>6</sup> The formulas for converting observed gross profit share into inferred gross profit share are:

<i>No-Shift</i>	<i>Full Wage Shift</i>	<i>Full Sales Tax Shift</i>
$w_k$	$w_k(1-\alpha)$	$\frac{w_k \cdot p - \beta}{p'}$

$w_k$  = observed gross profits including tax  $\div$  value of output in current prices.

$\alpha$  = corporate profits taxes  $\div$  observed gross property income including tax in current prices.

$\beta$  = corporate profits taxes  $\div$  real output in constant dollars.

$p$  = implicit price deflator for output = market price level of output.

$p' = p - \beta$  = inferred factor cost price level.

According to the originator of the method outlined above for measuring technical advance, if the relative increase in output per man-hour due to technical change  $z$  is unrelated to the capital-labor ratio  $k$ , technical change must be neutral, in the sense that the ratios of the marginal contributions to output of labor and capital are unaltered at any given capital-labor ratio. The empirical test for determining whether  $z$  and  $k$  are unrelated is the absence of any observed relationship between  $z$  and  $k$ , as revealed in a scatter diagram or regression. If no relationship is apparent, the presumption is that technical change over the observed period was neutral. This empirical test is not infallible, as its originator has acknowledged, but in the absence of information demonstrating the contrary, the test is presumptive evidence.<sup>7</sup> This empirical test indicated neutral technical change in manufacturing for the 1919-59 period on each shifting hypothesis.

Neutral technical change, along with marginal productivity pricing and constant returns to scale, permits the production relations between output per man-hour  $q$ , capital per man-hour  $k$ , and time  $t$ , to be expressed in the following form:

$$q = A(t) \cdot f(k)$$

(2)  $A(t)$  = index of technical progress at time  $t$ .  
 $f(k)$  = unspecified function of  $k$ .

With neutral progress, an index of technical progress  $A(t)$  can be constructed from the relative increase in output per man-hour due to the passage of time alone  $z$  (cf. equation (1)), since  $z$  is an approximate measure of the relative annual increase in  $A(t)$ . If the index of progress in the first year of a period is arbitrarily set at unity, and  $z$  for that year is  $z_1$ , the index of progress in the next year will be  $1+z_1$ ; and if  $z$  in that year is  $z_2$ , the progress index in the next year will be  $(1+z_1)(1+z_2)$ , and so on. By determining the value of  $A$  for each year through time in this way, the preliminary step in deriving a production function is achieved.

The next step is to derive output per man-hour deflated for technical change. Observed man-hour output  $Q/L=q$  for each year is divided by the progress index  $A$  for the corresponding year. The result represents an estimate of the output which would have been produced per man-hour with the capital endowment in use if "technology" had remained unchanged at the 1919 base level. This estimate of output with base year technology  $q/A$  depends upon the particular profits tax shifting assumption made for the calculation, since the progress index  $A$  is determined by the relative increase in hourly output due to progress  $z$ , and the latter reflects capital's inferred contribution to output  $s$ , as determined by the tax-shift hypothesis. With three alternative shifting

<sup>7</sup> Robert M. Solow, "Reply," *Rev. of Econ. and Statis.*, Nov., 1958, p. 413.



assumptions, there are, accordingly, three alternative estimates of deflated hourly output.

In principle, the data for determining the nature of the aggregate production relationship corresponding to each tax-shift assumption have been provided by these calculations. For each observation of estimated deflated output per man-hour  $Q/AL$ , there exists a corresponding estimate of capital used per man-hour  $K/L=k$ , and the relations between them can be observed in a scatter diagram or determined by simple regression techniques.

### *Consistency as Estimators*

Evaluation of the alternative tax shifting assumptions is accomplished (1) by comparing the closeness of fit of deflated output per man-hour against capital per man-hour and (2) by comparing how closely unde-  
flated output and property shares of income estimated from regressions

of the data and the correlated indices of technical progress are related to actual values. In these comparisons, simple functional relations are fitted by least squares to the data corresponding to alternative shifting hypotheses. These quantitative measures must be interpreted in the light of economic conditions, in weighing the evidence concerning tax incidence.

### *Evaluation*

If technical progress was on balance neutral over the 1919-59 period, it would appear that the traditional no-shift hypothesis yields more consistent results than the alternatives which postulate shifting. This is indicated by statistical and economic considerations. From a statistical point of view, the tax-shift model that offers the best fit between actual and estimated deflated output is the preferred variant, since the least squares requirement of minimum deviations between actual and estimated values applies only to the variables directly involved in the fitting procedure. The scatter diagrams between actual deflated hourly output and hourly capital indicated that several simple forms of functional relationships could be employed in the fitting procedure without altering the basic conclusions. Among these, two types were chosen to record the characteristics of the production relationship. The basic production relation fitted to deflated output, capital, and labor is a Cobb-Douglas function of the form

$$\frac{O}{L} = \left( \frac{K}{L} \right)^b, \text{ equivalent to } O = K^b \cdot L^{1-b},$$

where  $O$  is deflated output  $Q/A$ ,  $K$  represents capital employed,  $L$

stands for labor input, and  $b$  is a parameter indicating capital's contribution to output. This relationship was fitted by least squares to the logarithms of the variables for all years in the 1919-59 period. The basic linear regressions between logarithms of output per man-hour and capital per man-hour indicate the variance in deflated hourly output is 97.22 percent accounted for with the no-shift assumption, 94.55 percent with the sales shift, and 93.60 with the wage shift, for the capital variant including inventory. Approximately the same figures for explained variance are provided with the capital variant excluding inventory, the coefficient of determination  $R^2$  being .9715 for no-shift, .9432 for sales shift, and .9308 for wage shift.

The alternate production relation fitted to deflated hourly output and hourly capital is the straight linear form

$$\frac{O}{L} = c + b\left(\frac{K}{L}\right),$$

which is equivalent to  $0 = c \cdot L + bK$ , where  $c$  is a constant and the other symbols follow the designations previously described. This relationship was fitted by least squares to the actual values for the whole period. The results indicate that the no-shift model explains 96.07 percent of the variance in output per man-hour, the sales shift 91.58 percent, and the wage shift 90.24 percent, with the comprehensive capital definition. With the capital variant excluding inventory, the coefficient of determination is .9631 for no-shift, .9229 for sales shift, and .9081 for wage shift.

Other attributes of all regressions, basic or alternate, such as the standard error of estimate, also indicate that the no-shift and sales-shift models provide, respectively, the first and second best fits of production relationships, but there is evidence of serial correlation in the residuals between estimated and actual deflated hourly output. Though differences among the estimating errors of the three tax-shift models may appear to be small, they do record significant variations introduced by the shifting assumptions: the raw series utilized for each shift model were the same. The regressions for the whole period show that one tax-shift assumption yields the most unique or consistent production relationship for two different varieties of production function. Given the procedures of the method, these results are a strong argument for the priority of the no-shift case, and indeed, are largely sufficient as grounds for decision.<sup>8</sup>

However, other implications of these results are interesting and impor-

<sup>8</sup> Actually, no regression is needed with Solow's method to demonstrate the structural implications and the conclusion concerning tax shifting, since the structural distortions are evident in scatter diagrams of deflated hourly output and hourly capital.

tant on grounds of economic logic. On the a priori considerations that a unique production relationship exists with given technology and that diminishing returns to capital exist, comparisons of the inferred productivity of capital that are implicit in each shifting assumption suggest that the no-shift hypothesis is more compatible with diminishing returns to capital than the other hypotheses. In the derivation of deflated output, the method presumes that aggregate profits, adjusted for profits tax shifting, are the total return to capital as determined by marginal productivity imputation. Inferred marginal productivity of capital is thus total inferred contribution of capital divided by the capital stock. With an unchanged production relationship (in deflated output) over the period, the actual values of inferred marginal product of capital should indicate diminishing returns to capital as the capital stock increases relative to labor input, and the productivity of capital should be the same for equivalent capital-labor ratios. Chart I exhibits the scatter between inferred actual gross marginal product of capital and the capital-labor ratio on the no-shift assumption. The observations are reasonably close along a line indicating diminishing returns, with the exception of certain years of the Great Depression. Charts II and III show diminishing returns for the sales and wage-shift models, respectively, but a noticeable downward drift for observations in the 1942-59 period implies a sharp fall in the inferred productivity of capital. This decline in marginal productivity could not have occurred with neutral technical progress, since neutrality requires an unchanged production relationship.

Comparisons between actual output (or profits) and estimated output (or profits) derived from a fitted production relationship are interesting but hardly definitive in selecting a preferred tax-shift hypothesis. Such comparisons are interesting, because the regression results do not necessarily guarantee that the tax-shift models will estimate actual output (or profits) in the priority order established for deflated output as it appears in the regression. Conversion of deflated to actual output with the technical progress index—representing only one step in going from estimated to actual output—introduces by itself a variable not accounted for in the regressions. Though the fitted production relations implicitly determine capital's contribution to output, the closeness of fit insured applies to output rather than the property share. Such comparisons are not definitive, because it is difficult to choose between competing forms of production function with this method.

With the basic production relation—which displayed some advantages over the alternate linear form—the model that assumes no short-term profits tax shifting provides the most reliable estimates of actual (undeflated) output, and the wage-shift assumption the least reliable,

CHART I  
INFERRED MARGINAL PRODUCT OF CAPITAL WITH VARYING CAPITAL INTENSITY: NO-SHIFT

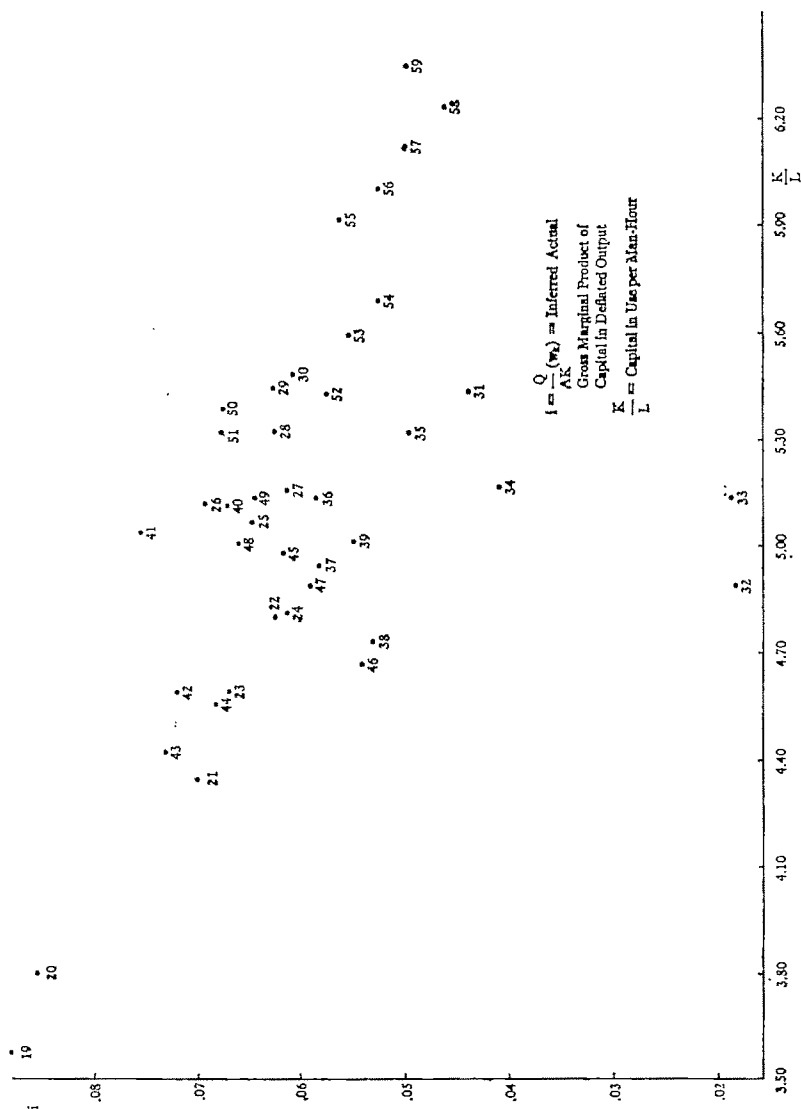


CHART II  
 INFERRED MARGINAL PRODUCT OF CAPITAL WITH VARYING CAPITAL  
 INTENSITY: SALES TAX SHIFT

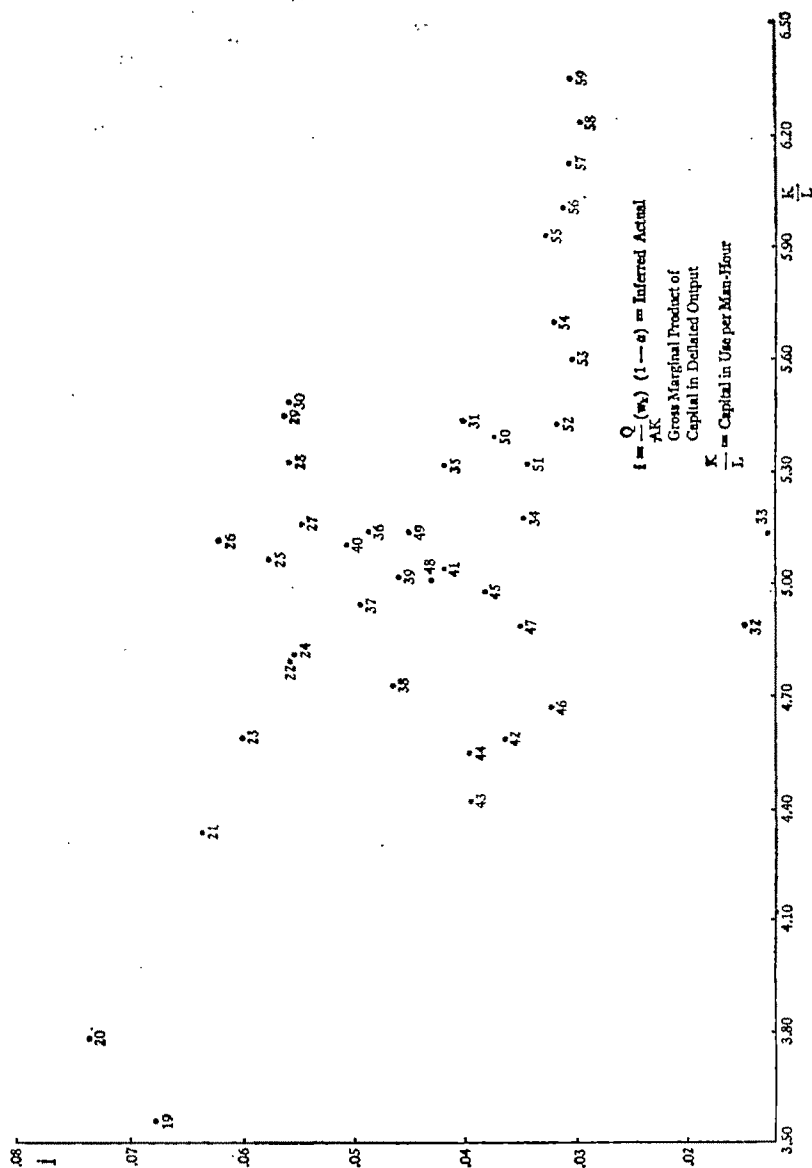
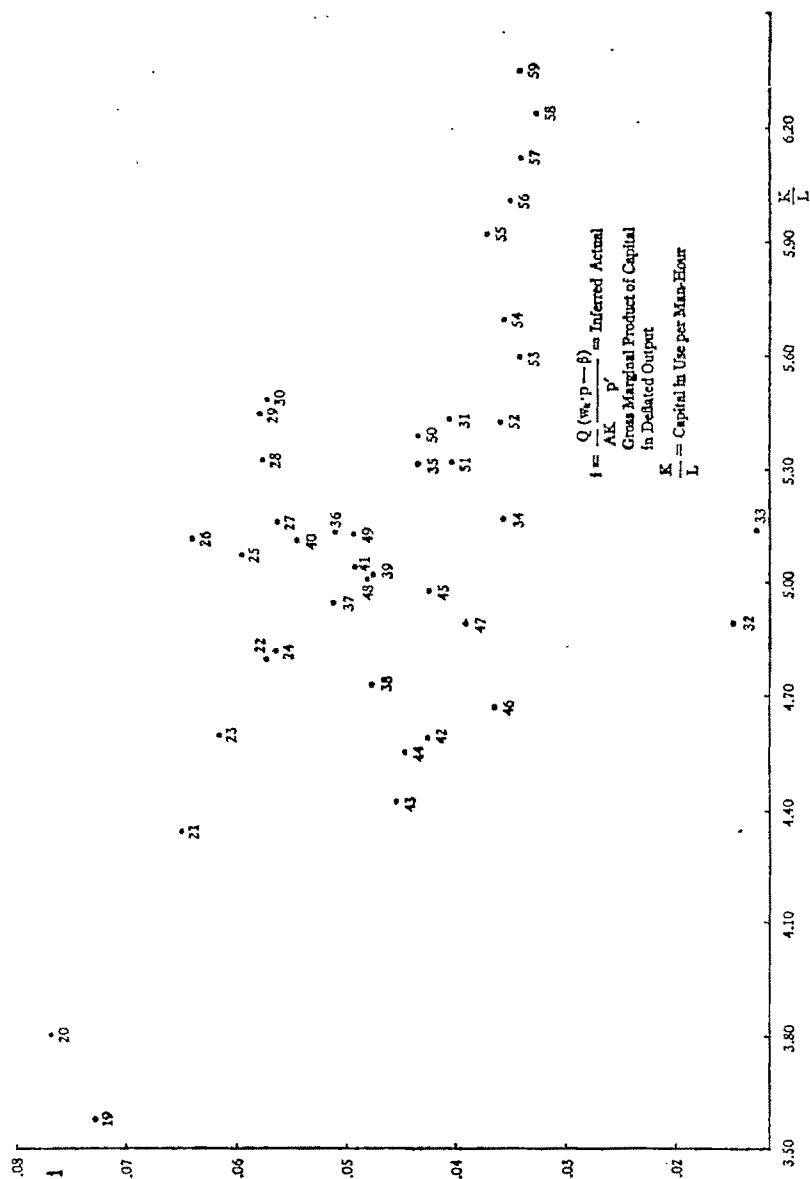


CHART III  
INFERRED MARGINAL PRODUCT OF CAPITAL WITH VARYING CAPITAL  
INTENSITY: WAGE SHIFT



as measured by the mean of the squared deviations of estimated from actual output. On the same criterion, the no-shift assumption provides the most reliable estimates of profits both before and after-tax in the models excluding inventory. But the sales tax-shift assumption yields more accurate estimates of profits before-tax than the no-shift hypothesis in the calculations including inventory in the capital variant, and the wage shift takes third place. On the other hand, profits after-taxes are most closely estimated with the no-shift model, and least closely on the wage shift, for both variants of capital, according to the least squares deviations measure. ✓

On balance the evidence therefore suggests that the traditional no-shift hypothesis yields results that are more consistent than the sales tax or wage-shift assumptions. The favored assumption yields a more consistent pattern of production relationship and inferred factor returns, and at least with the basic production relationship, estimates actual output more exactly, and, with one exception, predicts profits more accurately.

Nevertheless, it is not possible to reject with high confidence the hypothesis that profits taxes were directly shifted. This is attributable to the fact that the test for neutral technical change is not infallible, and reinforced by indications that tax shifting is compatible with non-neutral technical progress.

If technical change was significantly not neutral for any substantial interval over the 1919-59 period, the deflation of output per hour would not yield a consistent scatter between all observations of deflated hourly output and hourly capital; within the relevant subperiods, consistency should appear, but between relevant subperiods, the same capital per man-hour should show different corresponding deflated outputs. Moreover; on grounds of productivity imputation, the difference between the subperiods should indicate a change in the relative marginal contributions of capital and labor at any given level of capital intensity. The scatter diagrams of deflated return against capital per man-hour strikingly reveal these tendencies in the data. In Chart I (no-shift) deflated return at various capital intensities appears to be about the same for the 1919-41 and 1942-59 subperiods. However, in the shift cases (Charts II and III), deflated returns are evidently lower for equivalent capital-labor ratios in the later subperiod. Deflated hourly output against hourly capital (not shown) also reveal these tendencies.

Therefore, a hypothesis that profits taxation was shifted over the whole period is consistent with nonneutral technical progress, only if technical change reduced the productivity of capital relative to that of labor during roughly the last half of the whole period. If this relative decline in capital productivity did occur, in fact, the evidence would seem to indicate that the heavier profits tax burden of the 1940's and

1950's was indeed shifted.<sup>9</sup> In this event, there are no persuasive grounds favoring wage or sales tax behavior.

Nevertheless, the evidence available from this method suggests that technical change in manufacturing was generally neutral over the four decades and that profits taxation has not been shifted in the short run. }

\* The average annual ratio of federal and state corporate profits taxes to net profits before profits taxes increased from 15.89 percent for the years 1919-30 and 1936-40 to 51.00 percent in the period 1941-59. However, the annual average net profit rate (after taxes and capital consumption allowances) on the privately-owned gross stock of plant, equipment, and inventory was 5.94 percent in the earlier period and 5.97 percent in the later years of high tax rates. These figures are based on estimates prepared for the project.



## THE TAX STRUCTURE AND CORPORATE DIVIDEND POLICY\*

By JOHN A. BRITTAIN  
*Vanderbilt University*

This is a preliminary report on a model of corporate dividend policy which is designed to explain certain interesting features of dividend behavior since World War I. The first of these was a striking lag of aggregate net dividends<sup>1</sup> (relative to profits) from the late thirties through 1946. This was followed by a steady expansion at roughly a 6 percent annual rate<sup>2</sup> which tripled dividends by 1962. This postwar rate of growth of dividends was not far off the 8 percent pace of the 1919-29 period. What makes this recent dividend liberality especially interesting is that it occurred in the seemingly hostile environment of a "profit squeeze" which limited the trend in after-tax net earnings to a 2 percent annual gain. The postwar behavior also seems surprising in view of the continuing high individual income tax rates applying to dividend distributions. The model reported here suggests, however, that the profit squeeze is not entirely real. In part, at least, it is a statistical illusion engendered by liberalization of depreciation allowances; more realistic measures of corporate income show a more adequate economic basis for the postwar rise in dividends. Furthermore, the evidence produced by the model indicates that dividends (high as they may seem relative to profits) have indeed been substantially restrained by higher individual income taxation since the late thirties.

The focus of this study has been primarily on these two features of the tax structure—depreciation allowances and individual income tax rates—and their impact on dividends. However, insofar as changes in the corporate tax structure influence profits, the effect of this tax on dividends may also be estimated indirectly through its presumed depressing effect on after-tax earnings. The influence of these tax factors will be analyzed via statistical tests; since our interest is in fiscally induced changes over time rather than differences among firms and industries, the stress will be on time series regression models.

\* The material presented is taken from a current study supported by the National Committee on Government Finance and organized by the Brookings Institution. The writer is grateful to R. Goode for advice and criticism throughout, to C. Berry, R. Blitz, G. Fromm, J. Pechman, and N. Ture for comments, and to R. Rhomberg and B. Hickman for advice on computation. All are absolved of responsibility.

<sup>1</sup> Office of Business Economics estimates, after removing estimated capital gains of investment trusts and the "Rest of the World" component.

<sup>2</sup> Here and elsewhere "annual rate" of increase refers to the growth rate derived from a least squares trend line fitted to the logarithms of the observations.

I. *The Basic Lintner Model*

The approach of this study has been to generalize previously proposed models of dividend behavior by introduction of depreciation allowances, individual tax rates and other plausible explanatory variables. The underlying theoretical device relied upon repeatedly is the model presented by John Lintner to this annual meeting eight years ago. It relates aggregate dividends to lagged dividends and net profits (after depreciation and taxes). Writing net dividends in year  $t$  as  $D_t$ , net profits as  $P_t$ , and the unexplained random error term as  $u_t$ , the relation is:

$$D_t = a + b_1 P_t + b_2 D_{t-1} + u_t \quad (1)$$

The simplest rationale for this model is that dividends depend directly on current net income but are also constrained by past dividends, because of reluctance to cut dividends, or to raise them to higher levels which may not be maintained. The above model appears also in the work of Dobrovolsky and in the distributed lag formulation by Prais.<sup>3</sup> However, the particular pattern of corporate behavior suggested by Lintner is more amenable than the distributed lag to introduction of new variables.<sup>4</sup> The theoretical rationale also carries the endorsement of behavioral claims made by interviewed firms.

Lintner was led to relation (1) by the following model of corporate dividend policy:

$$D_t - D_{t-1} = a + c(rP_t - D_{t-1}) + u_t \quad (2)$$

Corporations were described as pursuing a "target" payout ratio  $r$  which they applied to current earnings  $P_t$ . If dividends were fully adjusted each year to achieve the target level  $rP_t$ , they would be changed by  $(rP_t - D_{t-1})$  between times  $t-1$  and  $t$ . However, a conservative bias against large revisions leads them to move only a fraction of the way, given by a "speed of adjustment coefficient"  $c$ . The constant term " $a$ " (expected to be positive) was added to allow for a presumed greater reluctance to cut dividends than to raise them. (If  $a$  is positive the rest of the model must call for a substantial downward adjustment in order for a cut in dividends to be indicated.)

One way of interpreting the Lintner theoretical framework is that dividends are determined by (1) capacity to distribute (indicated by  $P_t$ ) and (2) a long-run propensity to distribute ( $r$ ), modified by short-

<sup>3</sup> S. P. Dobrovolsky, *Corporate Income Retention, 1915-43*; S. J. Prais, "Dividend Policy and Income Appropriation," in *Studies in Company Finance* (B. Tew and R. F. Henderson, editors).

<sup>4</sup> Also relevant to the choice of model is the problem of autocorrelated error in the simplified distributed lag model, but methodological issues of this kind cannot be treated here.

run inertia or resistance. Although other types of models have also been tested, this general framework has proved the best vehicle for the present analysis, according to such standard criteria as goodness of fit and behavior of residuals. Also the treatment of short-run and long-run behavior within this model is convenient for the present problem. Since the primary objective is to account for the difference between prewar and postwar dividend policies, and the rapid postwar rise in dividends, a relatively long-run analysis is necessary. Otherwise it would be difficult, for example, to estimate the total influence of a cut in individual tax rates, because such a change might have only minor immediate effects.

Lintner's model provided an impressive explanation of interwar dividends, and the same fitted equation performed well when applied to the postwar period. However, some critics of the original formulation have stressed its aggregative character and its failure to explain the target ratio itself.<sup>5</sup> In his discussion of the model Tarshis also commented that the ability of this "relatively simple formulation to 'explain' dividends over a very long time span," during which "everything else in the economy changed" was a ground for suspicion of the results rather than satisfaction.<sup>6</sup> These criticisms can, however, be countered by an elaboration of the model. Rather than viewing Lintner's results as "a bit too good," as Tarshis did, it will be shown that they can be significantly improved by adding a propensity-to-distribute function and by taking account of some of the changes mentioned by Tarshis, including important shifts in the tax structure. The proposed treatment of the capacity and propensity to distribute will be taken up in turn in the next two sections.

## II. *Alternative Income Measures: The Problem of Varying Depreciation Allowances*

In Lintner's application of the model the current determinant of dividends (the source of funds from which they are paid) is taken to be net earnings (after taxes and capital consumption allowances)<sup>7</sup> as measured in the national income accounts. Interviewed firms had reported that net earnings was the key determinant of dividends, and it seems plausible as a measure of corporate capacity to pay dividends

<sup>5</sup> See, e.g., Jacob Michaelsen, "Determinants of Corporate Dividend Policy" (unpublished Ph.D. thesis, Univ. of Chicago, 1961). Some other criticisms by Michaelsen are less convincing. His own concern with interfirm differences leads him to dismiss time series models (such as Lintner's) as unimportant and uninteresting, mere "historical description." However, he does admit that his emphasis ignores their possible usefulness in explaining macroeconomic behavior.

<sup>6</sup> *A.E.R.*, May, 1956, p. 118.

<sup>7</sup> The series used was that estimated by the Office of Business Economics. The analysis was carried out first with the unadjusted profit series and then with profits including the inventory valuation adjustment.

during the interwar period to which Lintner applied his model. However, since the beginning of World War II five-year amortization and later liberalizing regulations and practices have clouded the meaning of tax return profit data.<sup>8</sup> This process of liberalization also distorts the regular profit series appearing in the national income accounts, since these rely upon tax return depreciation allowances. The proposition to be considered is that firms take into account these liberalized allowances and base dividend payments either on their total cash flow, or on some conception they may have of their "true" earnings.

*The Influence of Depreciation Allowances on Measured Profits.* The percentage of corporate profits (after tax) paid out each year in dividends shows some pronounced ups and downs since World War I, even apart from those induced by the greater cyclical variability of profits. This payout ratio climbed during the 1920's to an average of 72 percent in 1927-29, was cut to half that in the early postwar period, and has since risen to the level of the late 1920's.

First consider the postwar behavior of these series. While the dividend-net profit ratio  $D/P$  has doubled since the war, the ratio of dividends to "cash flow" (gross profits, net of taxes, but including depreciation allowances) has held remarkably stable near 30 percent since 1941. One hypothesis suggested by this contrast is that the postwar "profit squeeze" is in part a statistical illusion created by the depressing effect on measured net profits of rising depreciation allowances. Insofar as these increases are due to accelerated depreciation formulas or to increased administrative liberality on the matter of useful lives, the fall in tax return earnings simply saves taxes and can be shrugged off in deciding dividend payments. Indeed many firms use accelerated depreciation for tax purposes only, reporting the tax saving to shareholders as a liability for deferred income tax, or allowing it to "flow through" to increase reported earnings. Others allow the higher allowances to reduce profits reported to shareholders, but explain the noneconomic nature of the reduction.<sup>9</sup> In any case, liberalization imparts a downward bias to the tax return earnings from which profits are derived in the national income accounts.

The magnitude of this bias is difficult to assess, even sidestepping the controversy about how "true" depreciation should really be measured. However, there have been several published attempts to develop depre-

<sup>8</sup> A related tax factor which can also confuse the profit estimates is the expensing of research and development expenditures. A sharp rise in these expenditures since the war may have artificially depressed the net profits measure. Because this factor has been ignored below, the criticism of measured profits may be conservative.

<sup>9</sup> For a study of the contrast in earnings reported to stockholders and on tax returns see Edmund A. Mennis, "Different Measures of Profits," 1961 Proceedings of the Business and Economics Section, American Statistical Association, pp. 210-24. On the extent of use of accelerated depreciation see *The Federal Revenue System: Facts and Problems*, 1959, staff of the Joint Economic Committee, pp. 72-82.

ciation series that are at least consistent throughout and free of the influence of accounting changes. These series can be used for statistical estimation, without delving into the theoretical issue of which is "correct." Their value for the present study is simply that they abstract from variations in legal and administrative practice and provide alternative estimates of the breakdown of cash flow into two components—net profits and depreciation. Between 1946 and 1960 cash flow increased at an average rate of 6 percent per year. Several estimates of the break-

TABLE 1  
ANNUAL RATES OF CHANGE\* IN VARIOUS DEPRECIATION ESTIMATES FOR 1946-60, AND  
CORRESPONDING REVISED NET PROFITS SERIES

BASIS AND COVERAGE OF ESTIMATE	PERCENTAGE ANNUAL RATE OF INCREASE	
	Depreciation	Net Profits <sup>1</sup>
(1) Actual allowances, all corporations (OBE) . . .	12.9	2.0
(2) Original cost, straight line, Bulletin F lives, all corporations (OBE)† . . . . .	10.6	3.6
(3) "Realistic Depreciation," all business (MAPI)‡ . . . . .	8.2	4.1
(4) Actual allowances, all manufacturing corporations (OBE) . . . . .	13.1	2.4
(5) Original cost, straight line, Bulletin F lives, manufacturing establishments (OBE)§ . . . . .	8.3	4.6
(6) Same as (5), except current cost.§ . . . . .	8.8	4.1

\* Based on least squares trend lines fitted to logarithms of annual data.

† From data underlying Murray Brown, "Depreciation and Corporate Profits," *Survey of Cur. Bus.*, Oct., 1963.

‡ Derived from data on net stocks and capital expenditures underlying "Sixty Years of Business Capital Formation," George Terborgh (Machinery and Allied Products Institute, 1960). Reduced to corporate basis by means of estimates in Goldsmith, *A Study of Saving in the United States*.

§ *Survey of Cur. Bus.* and Supplements, adjusted to exclude unincorporated enterprise, but not converted from establishment basis to corporate basis.

down of this growth rate into rates of change in profits and depreciation are summarized in Table 1. In all series examined (including others not summarized here) the rise in estimated depreciation was considerably slower than the 13 percent annual increase in allowances reported in the national income accounts. Some estimates suggest that the latter allowances (permitted for tax purposes) overstate the rate of increase of depreciation by as much as 50 to 60 percent.

Attempts to adjust the MAPI "all business" series and the OBE establishment series for comparability with corporate allowances indicate that tax return depreciation allowances gained from about 55 percent of MAPI depreciation in the early postwar to almost 100 percent in 1960. Similarly, tax return allowances in manufacturing increased

from roughly 100 percent of the OBE original cost depreciation in manufacturing establishments to 150 percent by 1960. When these estimates are used to revise profits figures, a growth results which in some cases is double that of the official series, although still not matching the 5.8 percent growth of dividends. Nevertheless, this growth rate of dividends is equaled by that of cash flow which was 6 percent for manufacturing, as well as for all corporations. One hypothesis suggested by this pattern is that cash flow may influence dividends more strongly than either tax return profits or company estimates based on unchanging depreciation formulas. Such a dividend policy might seem irrational; yet, under recent circumstances, it may not be.

*Theoretical Rationale for Cash Flow as a Measure of Capacity to Pay Dividends.* Before presenting some regression models, a few hypotheses will be suggested that might explain any observed tendency of dividends to follow cash flow. They are (perhaps in order of increasing plausibility): (1) Firms may think of depreciation as a purely accounting charge not affecting their current ability to pay; cash flow (after taxes) then becomes a source of funds to be allocated between dividends and retentions without taking account of depreciation. (2) More plausibly, while recognizing depreciation as a cost, firms may regard stability or steady growth of dividends as a primary responsibility to be met before capital requirements. This could follow from the assumption that in practice dividends are a direct influence on stock prices, rather than a "mere detail," theoretically independent, as suggested by Modigliani-Miller. This might in turn lead to determination of dividends by cash flow, with any excess of gross investment (including replacement) over gross retentions to be financed out of excess liquidity, or externally. In the postwar period when most investment could be financed internally, the existence of such a policy would not be surprising. (Actually, even if capital requirements should have priority, the same link between dividends and cash flow could show up if the capital required tends to be proportional to cash flow.) (3) With the introduction of varying depreciation regulations it seems likely that firms would base their dividends on adjusted net profits utilizing consistent depreciation rules; however, for simplicity they may use cash flow as a proxy to represent the pattern of change in such an adjusted series. In other words, firms might desire to base dividends on true net profits, but assume that a better measure of trends in that variable is given by gross profits, from which, of course, a lower fraction would be paid out. This is analogous to the current tendency of national income specialists to reject the theoretically preferable "national income" series in favor of the GNP as a measure of economic growth.

*Statistical Tests of Alternative Explanatory Income Variables.* Estima-

tion of the model was based on thirty-eight annual observations for the period 1920-60, omitting the years 1936-38 to avoid the influence of the undistributed profits tax.<sup>10</sup> The timing of the two factors to be stressed in the dividend model suggested that the interval first be split into two nineteen-year halves. The 1942-60 interval includes the important changes in depreciation practices and is a period of generally high individual tax rates. This permits the depreciation issue to be studied in relative isolation, free of the disrupting influence of the sharp increase in individual tax rates that occurred earlier. (There were some additional rate increases up to 1944, and a strong upward trend in aggregate individual tax payments after that, due to higher incomes, but the main rate increases had occurred by 1942.) The same basic tests were applied to the more limited postwar period 1946-60 to check the possibility that the 1942-45 period was in some way peculiar. The resulting differences between 1942-60 and 1946-60 coefficients turned out to be insignificant; only the longer 1942-60 period is analyzed here.

The estimates have been made, not only for the corporate aggregate, but also for all manufacturing, and a study of individual industries and firms is in progress. Results for manufacturing closely parallel those for all corporations; only the latter are discussed here. In addition to tests of the net and gross profit variables  $P$  and  $C$ , two revised profit series were substituted for  $P_t$  in the Lintner model:

$P'$  = OBE net profits based on straight line depreciation at original cost and Bulletin F lives (variable 2, Table 1 above).

$P''$  = net profits adjusted according to an estimate of MAPI depreciation for all corporations (variable 3, Table 1 above).

Substituting a general income variable  $Y$  for  $P$  in relation (2) and multiplying out the right-hand side, the relations estimated were of the form:

$$D_t - D_{t-1} = a + b_1 Y_t + b_2 D_{t-1} + u_t \quad (3)$$

It seemed preferable to leave the model in first difference form as Lintner originally stated it in his A.E.A. paper rather than transpose the  $D_{t-1}$  to the right-hand side of the regression equation. Such a transposition yields the same regression coefficients, but the  $R^2$  measure approaches unity due to mutual trends in the independent and dependent variables. The first difference form reduces  $R^2$  but the result is probably a clearer measure of the variance that the model is explaining in its own right, without the aid of these mutual trends.

Least squares regression estimates for the four income variables are given in Table 2. Several features of these results point to cash flow as

<sup>10</sup> The substantial influence of this tax is interesting in its own right, since it imposed a penalty on retention along side the regular income tax penalty on distribution; however, no satisfactory way has been found to treat the two factors simultaneously in the model.

the best explanatory variable in this simple model, at least for the 1942-60 period. First, the substitution of any of the three alternative variables for  $P$  improves the statistical explanation of dividend behavior during this period of varying depreciation regulations. The substitution of cash flow does best, increasing  $R^2$  to .80, virtually halving the residual variance. This improvement in explained variance is easily significant at the 0.1 percent level<sup>11</sup> on the  $F$  test. The  $P'$  and  $P''$  models also provide better fits than the net profits variant, but estimates of the structural coefficients suggest additional reasons for regarding cash flow as

TABLE 2  
TESTS OF ALTERNATIVE INCOME VARIABLES ( $Y$ ) IN THE  
BASIC MODEL (3), 1942-60

$Y$	$a$ (billions of \$)	$b_1 = cr$	$b_2 = -c$	$R^2$	D-W*	$r = \text{TPR}^\dagger$
$P_t$	-.583 (.293)	.132 (.027)	-.147 (.045)	.610	1.79	.90
$P'_t$	-.472 (.228)	.162 (.025)	-.249 (.048)	.728	2.12	.65
$P''_t$	-.395 (.210)	.171 (.024)	-.200 (.039)	.757	1.94	.85
$C_t$	2.86 (.168)	.158 (.020)	-.540 (.074)	.801	1.90	.29

\* D-W is the serial correlation statistic used in the Durbin-Watson test.

†  $r = \text{TPR}$  is the "target payout ratio" estimated for variable  $Y$ , implied by equation.

the more influential determinant of dividends. The cash flow regression yields the positive constant term anticipated by a priori theory (and found by Lintner for the interwar period); also the target payout ratio of 29 percent is the level about which actual ratios fluctuated. The substantial negative constant terms<sup>12</sup> in the first three models and the high target ratios (especially the 90 percent value in the net profits model) have peculiar implications. It seems likely that the models are strained by the requirement to explain rapidly rising dividends by means of these sluggish profit estimates.

The inference that the relatively poor performance by net profits is due to the vagaries of the depreciation allowances since the war began is given added support by the fact that the same regression equations fitted to the 1920-41 period show all four income measures performing

<sup>11</sup> Here and elsewhere the term "significant at the 0.1 percent level" refers to the area in the tails outside the critical region of the distribution of the statistic.

<sup>12</sup> The constant term is expected to be positive for reasons discussed earlier. It should be recognized, however, that if the relationship is not linear, the fitted model could yield a negative constant term even if the true intercept were positive.



as well as cash flow for the later period; constant terms are positive, and other coefficients are plausible with  $R^2$  around 0.82.

*Other Issues in Evaluation of the Income Variable.* To supplement the tests of alternative profit measures, depreciation measures were explicitly included in the postwar model. For example, inclusion of allowances  $A$  along with  $P$  raised  $R^2$  to 0.81 and yielded a target payout ratio of 27 percent for  $A$ , only slightly below the 33 percent figure for  $P$ ; the coefficients for both variables were highly significant, with that for  $A$  positive and four times its standard error. (For the 1920–41 period, when the depreciation rate was less variable, use of  $A$  as a separate variable made no significant contribution.) Other tests with OBE and MAPI depreciation were also made, but in no case did the results improve on the simple cash flow model.

Another issue in the appraisal of the various income variables is the appropriateness of the inventory valuation adjustment (IVA). It seemed likely that before applying a payout ratio firms would adjust profits to allow for the changing cost of inventory replacement. To test this hypothesis a series of regressions was run on profit variables adjusted for inventory valuation. In all cases the adjustment yielded  $R^2$  substantially lower than those for unadjusted profits, and in many cases the Durbin-Watson serial correlation statistic was intolerably low. (For example, substitution of the adjusted series for  $P$  in the basic model fitted to the interwar manufacturing data reduced  $R^2$  from 0.8 to 0.6 and D-W from 1.71 to 0.89.) The IVA was then included as a separate variable with the same negative results. Some evidence was found that with a lag of one year the IVA was negatively correlated with dividends in the 1942–60 period, but not in 1920–41; the question deserves further investigation, but the IVA is not included in any of the models discussed here.

The net result of these tests, then, was that none of these profit and depreciation series could improve the fit yielded by  $C$  alone in the 1942–60 period. This does not prove that  $C$  is the best explanatory variable, because, e.g., it may only be acting as a proxy for some excluded variable. Still nothing has been found to date to undermine the cash flow as the best income variable in the 1942–60 interval. However, in analysis of the 1920–60 period it was found that the model was improved by explicit introduction of measures of depreciation. To sum up, the superiority of  $C$  in the later period could mean either that the two depreciation corrections utilized above were inadequate, or that firms were following cash flow in their dividend policy. The tests do not prove either of these hypotheses is correct, but the results are consistent with both.

### III. *A Propensity-to-Distribute Function: The Effect of Individual Tax Rates*

In view of the great changes since the thirties, especially in the tax structure, it would be surprising if the simple models discussed above could accurately describe the entire period since 1920. Lintner applied to 1946-54 his regression equation fitted to interwar data and found that it performed in this task better than various other dividend models. However, statistical tests suggest that (even apart from the issue of the best profit variable) the changes in relationships have been such that a more general model is needed.

*Evidence of Structural Change.* As a preliminary test the basic model (with net profits, unadjusted for IVA) was fitted to the 1920-41 data and applied to the 1946-60 interval. The standard error of estimate of \$695 million was only slightly larger than the figure of \$611 million obtained by Lintner when he applied his interwar (1918-41) regression to the 1946-54 period. These errors are small relative to total dividends, as Lintner pointed out, but by other criteria they are large. They are double the standard error of estimate of the same model fitted to the 1942-60 period, and triple the standard error of the cash flow model fitted to that same period. The postwar annual increase estimated by the interwar equation is too large every year and averages over \$1 billion, or double the actual average.<sup>13</sup>

To supplement these descriptive comments, a standard analysis of covariance test was applied. If the goodness of fit for the model estimated separately for the first and second halves is significantly better than the overall estimate for 1920-60, some kind of structural change is indicated. For the basic model the standard errors of estimate were \$262 million and \$308 million for the first and second halves and \$355 million for the entire period. The discrepancy is significant at the 1 percent level on the  $F$  test. The cash flow model fits the entire interval better than the net profits model ( $R^2=0.702$  and  $0.637$  respectively; level of significance=1 percent), but it too fails the structural change test, as do the other two profit variables tested. The same is true when depreciation allowances are included along with net profits; this makes no improvement on the cash flow model. Another indication that structural change is plaguing all of these models when fitted to the entire forty-year interval is the indicated auto-correlation of the residuals (with  $D-W$  values at 1.3-1.4 and a tendency to underestimate prewar and overestimate later). Exclusion of a key variable could cause this.

The above tests tend to support the original comment of Tarshis

<sup>13</sup> Lintner points out that the profits variable including the IVA does better on this latter count, but it performs about the same according to the standard error of estimate and other criteria, and less satisfactorily according to the  $R^2$  and  $D-W$ .

about the oversimplicity of the basic model containing only profits and lagged dividends as independent variables. Different structural relations appear to be present in the two time periods. One of the most likely elements in this structural change is the fact that the first period was one of relatively low taxes, while the 1942-60 interval was one of generally high taxes. Since an attempt has already been made to allow for corporate taxes and depreciation allowances by the use of after-tax cash flow, the next factor to consider is the individual income tax rate structure.

*A Hypothesis Relating Payout Ratios to Individual Tax Rates.* The payout ratio may tend to vary inversely with the differential between tax rates on ordinary income and capital gains. Assuming shareholders do not generally have their hearts set on any particular after-tax dividend rate, this association would seem likely in view of the tax-saving (or shelter effect) afforded by corporate retention. The behavioral rationale underlying this hypothesis is that the payout ratio adopted by boards of directors is subject to one or more of the following influences: (a) the personal tax-saving possibilities of high-income owners may be recognized and directly pursued by boards (especially in closely held companies) who seek to shelter owners (including themselves) from taxes; (b) on a more sophisticated level, firms may be trying to maximize the discounted value of all after-tax future returns and recognize that a rise in personal taxes may lower the optimum payout ratio; (c) the preferences of high-income (capital gains-preferring) shareholders, as expressed in the capital market by their relative demand for securities with different payout ratios may move boards of directors to let payout lag when tax rates rise.

*Derivation and Measurement of the Shelter Variable.* In order to test the association between dividends and tax rates, it is necessary to develop and build into the model some kind of measure of individual rates. Within the theoretical framework of (b) above, and assuming equal tax rates for all shareholders, Myron Gordon derived an optimum payout ratio inversely dependent upon  $t_y/t_o$ , where  $t_y$  and  $t_o$  are the tax rates on ordinary income and capital gains, respectively.<sup>14</sup> From the point of view of the behavior mentioned in (c) above, it is appropriate instead to measure shareholder incentive to hold low payout securities. Measures of this "shelter effect" that can be derived under such assumptions are  $(t_y - t_o)$  and  $(1 - t_o)/(1 - t_y)$ .

A second methodological issue is involved in the measurement of  $t_y$  and  $t_o$ . The weighted average marginal rate on dividend income seemed the most appropriate variable, but for simplicity, and also because of the possible special influence of high-income groups, an alternative was adopted. Several annual income series were developed for 1920-60

<sup>14</sup> *The Investment, Financing and Valuation of the Corporation*, pp. 131-39.

which cut off at the top of the income scale a given fraction of all dividends. The marginal values of  $t_v$  and  $t_o$  for certain family structures were then estimated, and the resulting time series were used as explanatory variables.

*The Model and Some Statistical Results.* The relationship suggested is between the (long-run) target payout ratio and the shelter variable (the tax incentive to hold low payout securities). The target function was therefore added to the single equation payout model. For example, taking  $(t_v - t_o)$  as one measure of the shelter effect, the simple linear relation that suggests itself is:

$$r_t = \alpha + \beta(t_v - t_o). \quad (4)$$

This relationship can then be substituted for the constant  $r$  in the basic model, (3) above, to test the hypothesis that  $\beta$  is negative. Various tests produced the interesting result that  $t_o$  was not essential and made no contribution, even when included separately. (This could reflect the stability of  $t_o$  relative to  $t_v$ , rather than irrelevance of  $t_o$ .) Since the different measures of  $t_v$  were adequate by themselves, a sample of regressions is given taking  $t_v$  alone as the shelter variable; one particular variant which measures the marginal rate for a married couple with two dependents at income levels cutting off the top 10 percent of dividends will be used as an illustration, measured in decimal form and labeled  $t_{10}$ . This particular index of individual tax rates in high-income brackets dropped from 0.56 in 1920 to 0.20 during 1925-31, rose to a peak of 0.92 in 1944-45 and has been within the range of 0.74-0.85 since the war. With the introduction of the target function cash flow continues to give the highest  $R^2$ , and this regression is given in (5) below (as estimated for 1920-60, excluding 1936-38). The inclusion of the shelter variable reduces the unexplained variance by 40 percent, and its coefficient differs from zero in the expected direction by five standard errors.

$$D_t - D_{t-1} = .474 + .198 C_t - .097 t_{10}C_t - .430 D_{t-1} \quad (5)$$

(.112) (.020) (.019) (.042)

$$R^2 = .829 \quad D-W = 1.75 \quad r_t = .460 - .224 t_{10}$$

The standard error of estimate is reduced 25 percent to \$243 million. More interesting is the fact that this standard error of estimate is only slightly inferior to those for separate fits of the same model to the first and second time periods, which are 230 and 220, respectively. The covariance test yields a remarkably low  $F$  ratio of 1.26, showing no trace of structural change now that the shelter variable has been built into the model. This does not, of course, prove there has been no such change or that the model is correct, but it is interesting to find that the strong evidence of its existence in the simple model disappears when individual tax rates are taken into account.

Regression equation (5) implies the target function  $r_t$ , which is given below the equation. If this relation continues to be indicated when other variables are added, insertion of the  $t_{10}$  values would suggest that individual tax rates had depressed the target  $D/C$  ratio about 35–40 percent between the late 1920's and the end of World War II. The estimated ratio dropped from 41 percent to 26 percent; the ratio was still down about 30 percent by 1960, implying that the target level of dividends was about \$5–6 billion below what it would have been at the tax rates of 1925–30. Actually, however, some evidence was found that for the entire period the model with cash flow and the shelter variable could be improved by inclusion of explicit depreciation variables.

*The Combined Influence of Allowances and Tax Rates.* To estimate the effect of changing allowances and individual rates simultaneously both were included in the model in another series of tests. In some cases the depreciation variable was simply added to the regression; in others it was included within the target model. The latter assumes that firms have in mind some measure of true depreciation and profits when deciding their long-run dividend policy. The result of one such test is given in (6). The target model includes the ratio  $A/A''$  of actual allowances to MAPI depreciation as an index of the liberality of the allowances. The term containing the depreciation index proved highly significant; as shown in (6) the target ratio  $r_t$  is positively associated with  $A/A''$ . This estimated target ratio is the continuous plot in Chart I.

$$D_t - D_{t-1} = .249 + .125P_t + .109 A/A''P_t - .105t_{10}P_t - .239D_{t-1} \\ (.122) \quad (.034) \quad (.030) \quad (.029) \quad (.032) \quad (6)$$

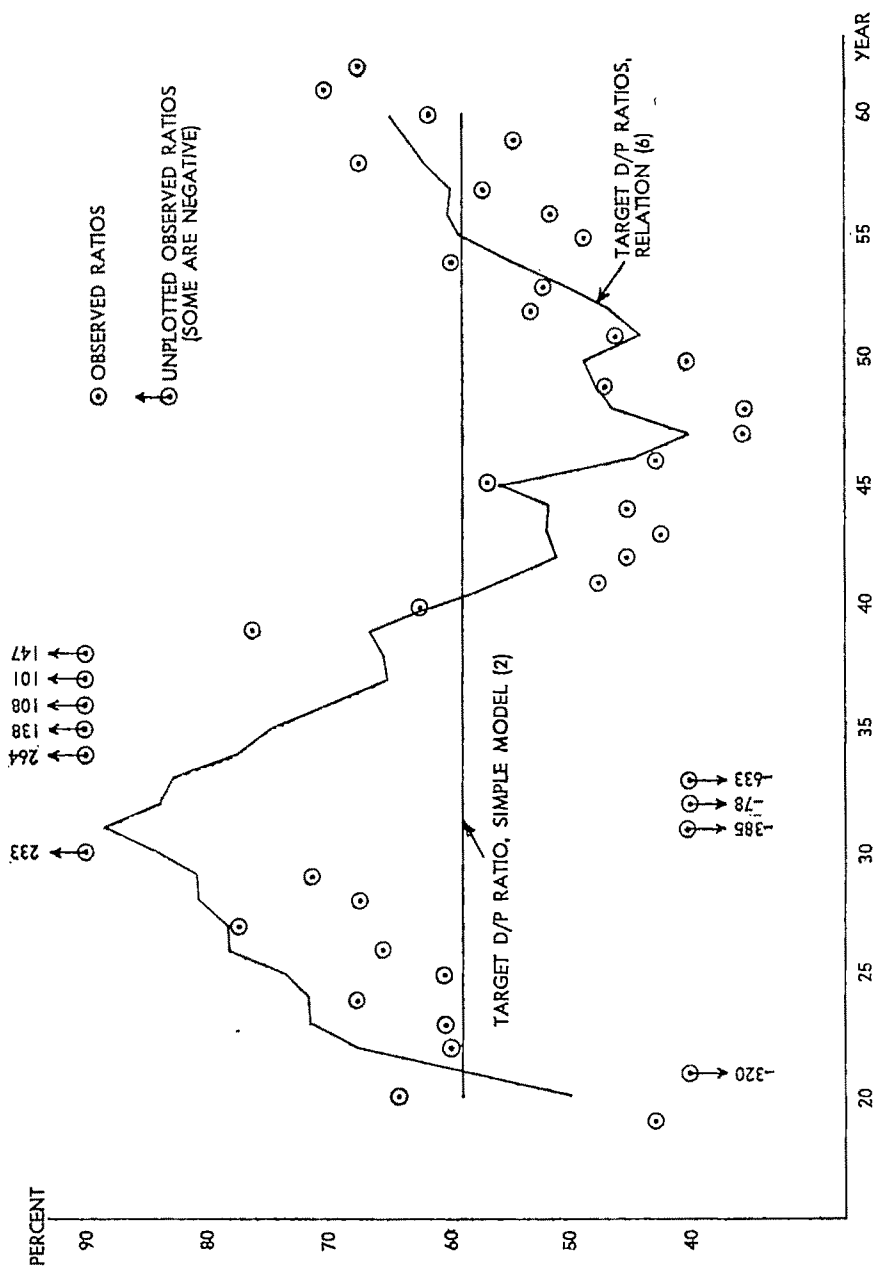
$$R^2 = .819 \quad D-W = 1.85 \quad r_t = .524 + .456A/A'' - .441 t_{10}$$

The observed ratios depart substantially from the target in depression years; the denominator approaches and even passes through zero as profits became negative in four years. Otherwise, the target function (based only on depreciation and individual rates) explains well the long-run swings in the payout ratio—the rise during the 1920's (mainly falling tax rates), the fall between the late 1920's and the war (mainly rising tax rates), and the rise after the war (mainly increased liberality of depreciation). This variable target ratio may be compared to the horizontal line on the Chart which represents the constant ratio estimated by the original Lintner model.

#### IV. Other Factors in Dividend Policy

The tentative findings discussed above have been derived from an elaboration of Lintner's model in which only tax variables have been added. Tests of other hypotheses are essential if the results are to become more convincing. One such hypothesis is that the low postwar

CHART I  
OBSERVED AND TARGET D/P RATIOS



payout ratio was the result of a heavy investment demand which could be financed most cheaply from internal funds. Investment variables such as new plant and equipment expenditures (and the same variable with inventory change added) have been tested within the framework of this model with negative results, which supports Lintner's earlier findings. If true, this suggests that funds are allocated to dividends on the basis of income, with modification by tax factors, and without regard for the quantity of capital which must be raised externally. This tentative result is certainly interesting but also rather surprising. It is possible that with disaggregation of the model (so that fast growing industries and firms can be studied individually) some indication may appear that rising investment depresses dividends.

The effects of corporate liquidity and working capital were also considered, and the former factor appears to have a positive (but statistically marginal) influence on dividends. (In no case so far has inclusion of this or any other variable undermined the explanatory influence of the two tax factors.) One other factor which appears correlated with dividends is the change in sales, especially the two-year change suggested by Darling.<sup>15</sup> His argument was that this variable would give an indication of anticipated working capital needs. The sales change does appear to have a negative influence on dividends, but in absence of any influence of investment demand when studied explicitly in our earlier tests, the interpretation of this is difficult. Perhaps a more likely rationale is that a rapid sales increase is a proxy for a rapid gain in earnings (not tested directly) which firms may feel will not be maintained. This rate-of-change factor could encourage a lag in payout in addition to that described by the speed of adjustment coefficient  $c$ , or it could affect  $c$  itself.

Another area which deserves consideration is the possible relationship of dividends and stock prices. Without becoming involved in the extensive theoretical controversy on this question, it might be of interest to explore the possible feedback effects of price changes on dividends. If, e.g., a dividend lag does actually depress stock prices, the recognition of this effect might lead to a compensatory upward adjustment of dividends.

### V. *Tentative Conclusions*

Due to the exaggeration of the profit squeeze by liberalization of allowances, the cash flow model provides a better explanation of dividends since the interwar period. Introduction of a target payout model dependent upon individual tax rates and a depreciation variable gives a satisfactory explanation of dividends in the entire 1920-60 interval.

<sup>15</sup> Paul G. Darling, "The Influence of Expectations and Liquidity on Dividend Policy," *J.P.E.*, June, 1957, pp. 209-24.

These two tax factors appear sufficient to account for the sharp drop in the payout ratio between the late 1920's and the early postwar period, and its subsequent recovery. If corporate taxes are also taken into account via their effect on after-tax income, the net effect of the tax structure on dividends appears substantial. While the by-products of these changes in dividends cannot be discussed here, one of the most interesting is the effect on income distribution; by depressing dividends the tax structure has played an important role in the leveling of pre-tax personal incomes since the 1920's.<sup>16</sup>

These conclusions on the effects of the tax structure on dividends are tentative, especially since they are based on the corporate aggregate. While results for the manufacturing sector (about one-half the total) confirm the aggregate results, the study (now being carried out) of individual industries and firms may show substantial variations in behavior among different sectors. These individual firms and industries are also being examined for cross-section evidence on the main behavioral hypotheses of Section II and Section III.

<sup>16</sup> This might not be true of the *Statistics of Income* distributions, since the reduction of dividends could be offset by a corresponding increase in capital gains. However, no such offset can occur in the more frequently cited OBE estimates, since they exclude capital gains.



## ECONOMIC EFFECTS OF THE CAPITAL GAINS TAX\*

By MARTIN DAVID  
*University of Wisconsin*

An appraisal of capital gains taxation raises formidable problems. To know how the tax affects our economy we need to know how the tax affects each and every saver in his investment decisions. This requires an understanding of the investor's reaction to risky and uncertain situations, a model of the manner in which expectations for the future are built on limited and sketchy knowledge of the present, and an appraisal of the costs of making investments and investment decisions. Indeed the economic effects of the capital gains tax go further. Preferential taxation of capital gains may affect work effort, resource allocation, the total volume of savings, the structure of industrial organization, and other significant aspects of our economy.

What are the economic effects of capital gains taxation? The discussion here presents three aspects of these economic effects that are of major policy importance. I shall address myself to the impact of capital gains taxation on economic growth, the impact of capital gains taxation on stabilization of the economy at full employment, and the impact of the tax on efficient use of our economic resources.

Before we proceed, it is necessary to define these terms more precisely. I use capital gains taxation to refer to the past and present U. S. system of taxing income acquired through a variety of transactions in real and financial assets. With the exception of the years from 1917 to 1921, this system has had four principal features:

1. The capital gains tax has been applicable to a generic classification of transactions; the boundaries and principles establishing boundaries for the classification have been uncertain and arbitrary [16].

2. Holding the asset for given minimum periods has been one of the most important criteria delimiting favored transactions. The capital gains tax applies differential rates to income earned on favored transactions; the nature of the differential has varied from time to time and has differed for personal and corporate entities.

3. The capital gains tax has not been applicable to income except when realized in an actual transaction; this has meant that gains ac-

\* This paper is an outgrowth of preparation for a research conference on capital gains taxation. The conference is one of a series undertaken by the National Committee of Government Finance of the Brookings Institution in connection with its program of Studies in Government Finance which are financed by a special grant from the Ford Foundation.

crued on assets subject to gratuitous transfer have been exempt from tax under income tax laws; tax on the gains accruing on securities held over a long period of time have been applied only after considerable postponement [14].

4. Lastly, the tax law has limited the extent to which losses on transactions in eligible asset can be deducted from other forms of income; again the form of loss deduction and its amount have varied during the history of this law [3].

In discussing effects of capital gains taxation I shall refer to the differential behavior of the economy under two tax systems: (1) a tax system yielding revenues equal to those collected under the present system but without the preferential features for asset transactions<sup>1</sup> and (2) a tax system such as is actually in effect. The comparison does not refer to the difference between the present tax structure and the same structure with preferential capital gains tax provisions deleted because a closing of all capital gains loopholes would generate different amounts of revenue. The desirability of alternative levels of revenue and their economic effects are issues that need not be discussed in this paper.

### *I. Capital Gains and Economic Growth*

Many arguments have been voiced to defend preferential capital gains taxation. The argument that concerns us most asserts that preferential capital gains taxation contributes significantly to the economic growth of our national product and its quality. The argument is seldom spelled out in full, but it appears to me that the many complaints that taxation of capital gains locks-in the investor, that payment of such taxes draws funds away from the capital markets, that preferential taxes are required to induce risk-taking, and that preferential capital gains taxes are needed for a healthy capital market imply the following economic argument [2] [5] [10] [13] [15].

Preferential taxation of capital gains reduces the effective tax rate on income saved in the form of capital assets. On balance this discriminatory subsidy will favor saving in the form of assets eligible for the low tax rate. As a consequence (1) more funds become available in the capital markets; (2) investors respond to the increase in the supply of funds with increased outlays on real capital; (3) increased outlays raise the level of capacity production and may increase the rate of de-

<sup>1</sup>Theoretically it will make some difference to our conclusions as to how the burden of taxation is distributed in the uniform tax system proposed; we assume that progression in the uniform system can be made to correspond to a pattern not too dissimilar to the present rate structure on ordinary income albeit at lower rates. The change to full taxation of gains thus implies a decrease in taxation for individuals that receive no capital gains and an increase in taxation for individuals that receive a sufficiently large portion of their income in capital gains.

velopment of new technology and its adoption; (4) the economy benefits from preferential capital gains taxation by increased economic growth.

While this line of argument may satisfy a few economists who do not choose to look beyond a partial equilibrium analysis, I question its significance. What quantitative effects on growth can we attribute to preferential taxation? What benefits do we get from this indirect tax device as compared to other actions to stimulate growth? Let us examine the links in the chain of argument one by one.

First, capital gains taxation increases the volume of savings as it discriminates in favor of savings accumulated in assets privileged to the low tax. To evaluate this argument we must recall that this tax provides for far more than a marginal change in rate of return. The tax has significant income as well as substitution effects. The taxes that investors avoid by the capital gains loophole reduce both the average rate of taxation and the marginal rate applicable to capital gains. Households will be sensitive to the difference. They will adjust work effort, consumption, and saving in noneligible forms of investment in response to adjustments in their income. What clues do we have as to the outcome? Break suggests that for high-income groups the effective tax rate may have little to do with work effort [4]. But that analysis speaks for the behavior of a limited group of professional types; we can only guess about the response of lower-echelon corporate executives and owners of small businesses to lower effective tax rates. Do they retire early, turn down post-retirement jobs, or work shorter hours because of increases in disposable income?

Let us assume, on balance, adjustments in work effort for investors are unimportant. What else can be said about investor's behavior? If work effort remains constant, lower average taxes mean more disposable income for the investor in capital gains; in addition low marginal taxes on additional earnings in the form of capital gains will influence the investor's choices. Greater disposable income may be entirely consumed, but that is most unlikely. Preferential capital gains taxation implies greater savings among investors, provided that saving responds elastically to increases in yield. (While this seems likely, if target saving occurs on a sufficiently wide scale, net saving may be interest inelastic.) The corollary of increased savings in eligible assets surely is relatively less savings in ineligible forms. Relatively fewer dollars are invested in savings accounts, while more dollars are invested in mutual funds and stocks.)

In addition to changing the volume of savings capital gains taxation will also have an effect on the portfolio that the saver demands. Changeover from uniform income taxation to a preferential tax for

capital gains implies: (1) a lower average rate of taxation and (2) a reduction in the loss offset permitted. Following the arguments of Domar and Musgrave [7] and Musgrave [23] we must conclude (1) a lower average tax reduces risk-taking, and (2) limitations on loss-offset reduce risk-taking. If this model is adequate, a capital gains tax probably does not increase total demand for risky assets, or total risk-taking in the economy; i.e., income effects probably exceed substitution effects. Rather, the reduction in government participation in losses more than offsets the increase in private investor risk-taking. We note that present limitations on loss offsets would be unnecessary in a system of uniform taxation. They are made necessary by the realization criterion for determining the timing of tax payments.

One further point must be raised. In a full employment economy each dollar of tax subsidy favoring capital asset transactions must be matched by a dollar of additional tax yields from other revenue sources. Thus higher tax rates for noninvestors must be considered as part of the general equilibrium. We have little evidence as to how noninvestors might adjust their behavior, but it is possible that their work effort is more sensitive to tax rates than the work effort of investors. The assembly line worker, the clerk, and the laborer generally have fewer nonmonetary benefits and derive less satisfaction from the performance of their jobs than do the self-employed, the executive, and the professional. In addition, job satisfaction among less skilled workers hinges more heavily on the rate of take-home pay for less skilled workers than for executives and professionals [8]. Studies of such groups indicate increases in absenteeism and job turnover with decreasing job satisfaction. Thus under a system of capital gains taxation the losses associated with work disincentives for the average noninvesting worker may be considerably greater than the gains of additional investor savings [8] [17].

Let us return to the main thread of the argument. The second point stated capital gains taxes increase the supply of savings (and, as I have noted, may decrease aggregate demand for risky opportunities). What impact will the increased supply have upon investment? An increased supply of funds will mean more offers to buy both new and existing securities, both new and existing assets. The investor is indifferent between the new issue of an old company and outstanding equities. What becomes of additional funds supplied to the capital market thus depends on the demand for new investment funds relative to the available supply. If there is no demand for investment funds above and beyond the amounts that business can obtain from internal sources such as depreciation and the retention of earnings, the entire increase in supply of savings becomes translated into an

increase in the price and value of existing securities. If the available supply of funds exceeds demand for new investments, prices rise but by less than in this extreme case. Conversely, if the demand for new investment funds exceeds available supplies, equity prices fall and divert a part of existing asset value into new investment uses.

It appears that the ultimate impact of additional savings generated by preferential capital gains taxation on investment can only be evaluated by a model that includes investment demand. Despite the work of Jorgenson [11], Meyer and Kuh [20], and many others, our knowledge of the investment function is extremely incomplete. We do not know whether financial factors such as the availability of funds under conditions that will not precipitate declines in stock prices are an important or dominant element in investment demand. The anticipated rate of consumer spending and spending for the national product would appear to be more important factors. When spending rates are high, plant operates near capacity. Demand for investments is created by physical needs as well as the anticipated success of higher quality products and more productive techniques. At the same time the high rate of spending generates investor optimism that contributes additional funds to the available supply. Business activity results in a high level of profits and correspondingly high retained earnings. The overall effect in the last decade has been that corporate needs for financing new investments by new issues of securities have not been strongly influenced by the volume of investment in new plant and equipment. In fact, in the recession troughs of 1958 and 1961 corporate business financed a higher proportion of new investment from new issues than in any other year but 1957.

If it is uncertain that business has large marginal needs for financing additional investments during period of peak activity and if during periods of slack our major economic need is for larger spending plans rather than greater savings, where does the additional supply of funds for investment that is generated by preferential capital gains taxation really pay off?

The third point in the chain of reasoning that stretches between capital gains taxation and economic growth stresses a nexus between investment in plant and equipment and the increase in capacity GNP. Available studies indicate that investment in construction, machinery, and other producer durable goods plays a minor role in economic growth. Massell [18] demonstrates that 90 percent of the growth in output per man-hour of labor from 1919 to 1955 can be accounted for by technological factors, information, organization and similar factors. Only 10 percent is accounted for by increases in capital stock per worker. Similarly Denison [6] estimates that it would take an increase

in private investment of 1.4 percent of GNP to raise the rate of growth by even 0.1 percentage points. If greater availability of funds could be credited with reducing the lag between the discovery and adoption of new techniques by two and two-thirds years, the availability of funds could be credited with improving the rate of growth by another 0.1 percentage points. Financial devices to reduce the cost of capital appear to have little impact on growth relative to these requirements.

What about the arguments that full taxation of capital gains locks investors into undesired portfolios? Keeps entrepreneurs from investing in risky enterprises because their funds are tied? What of the trickle-down theory that asserts new funds are placed largely in gilt-edged securities so that old investments must be liquidated before additional risks are taken? These arguments highlight the fact that the combination of full taxation of capital gains and payment of tax at realization convert taxation from an involuntary payment into a postponable transactions cost. The greater the transactions cost, the greater the impediment to trading in the capital markets. ~~Two observations seem to be in order:~~

First, to the extent that capital gains are taxed on an accrual basis, no lock-in is possible. For nearly half of eligible transactions, in terms of dollar volume, accrual taxation of capital gains could easily be accomplished by a year-end accounting no more difficult than an inventory valuation. Publicly-traded corporate securities could be valued on the basis of well-established market prices.<sup>2</sup> Then the investor is unable to postpone tax payments and can trade without being locked in. The difficulties in valuing investments that are not readily salable has lead many to believe that accrual taxation is impractical and a prohibitive enforcement problem. Two options exist that would deal with the problem: First, the taxpayer could be allowed the privilege of paying his tax on realization if he also is willing to pay interest on the funds he acquires through the deferment of taxes. This could be accomplished in a rough way by taxing gains as if they accrued at a constant rate over the period held. Alternatively, Turvey [24] proposes that valuation of closely held investments be avoided by direct taxation of retained earnings to the individual; this procedure is already available to those corporations that elect to be taxed under Subchapter S. The procedure might be modified for other corporations that elect to allocate retained earnings to shareholders for tax purposes in lieu of valuing equity holdings.

<sup>2</sup> *Statistics of Income: Supplemental Report, Sales of Capital Assets Reported on Individual Taxes for 1959*, Table 2 indicates 50.1 percent of all realized net long-term capital gains were taken on corporation stocks, U.S. obligations, and other bonds. Some of these are undoubtedly close corporations but the overwhelming majority could be correctly valued.

Second, adjustment of portfolios to meet the demands for funds from new investment undertakings is required only if the risk characteristics of new investment opportunities are distributed differently from risk preferences of current savers. If the supply of funds is directed towards somewhat less risky holdings than the opportunities which new investment affords, the market may still reach an equilibrium without major changes in the relative prices of high- and low-risk securities. This comes about as some existing securities change risk status with increasing maturity. The risk attaching to a security will decline with age. Old undertakings have survived the market test and are less likely to exhibit the unstable growth of new corporations. As a consequence, some investors must exchange maturing securities for new securities in order to maintain their desired level of risk. This turnover will accommodate certain discrepancies between the risk preference of savers and the riskiness of new investment opportunities.

Transactions costs will limit the extent to which this roll-over of investments accommodates differences in the character of supply and demand for funds. However, we do not know the importance of such accommodations relative to the volume of new funds reaching the capital market. Without some quantitative indications it is difficult to evaluate the benefits of capital gains taxation in this context. Furthermore, it does not seem appropriate to accommodate differences in supply and demand, not in proportion to the discrepancy that exists, but rather in proportion to the degree of realization of past economic gains in the form of capital gains. In effect, preferential capital gains taxation provides a tax subsidy to holders of existing assets in proportion to income earned on those assets. This subsidy does not necessarily vary in phase with the need for additional funds to finance new investments of a risky nature.

## II. *Capital Gains and Stabilization*

Of all the tax provisions in our arsenal of defenses against recession and inflation, preferential capital gains taxation is perhaps the most contrary. Taxation at the time of realization makes it possible for investors to offset cyclical fluctuations in their tax rate. Thus tax yields excluding capital gains taxes are more responsive to changes in personal income than changes in tax yield as a whole [12].

In addition, taxation at the time of realization limits the stabilizing power of progressive tax rates for high-income taxpayers whose actions are more likely than the average to precipitate marked changes in investment and consumer spending. If preferential capital gains treatment of income continues, we can expect this limitation to increase in its significance for the economy. The proposed 1963 tax reforms call for a reduction in marginal tax rates for all brackets. As

a consequence, individuals will experience less fluctuation in the marginal tax bite as incomes fluctuate in future years. Spending behavior will be proportionately less responsive and the stabilizing power of the income tax reduced from its present level. }

Investors' ability to offset stabilizing changes in their tax payments is aggravated by the fact that stock prices lead fluctuations in personal income.

If capital gains on corporate securities were taxed in full on accrual basis, cyclical variation in stock prices would produce about perhaps ten times as much response in the corresponding tax yields as the present system [1]. Response on this order could provide an extremely useful damper on destabilizing speculative movements in the market if full taxation also reduces the lead time between asset price movements and changes in personal income. If the lead pattern is not altered, full taxation of capital gains would be more destabilizing than under the present system.

### III. *Capital Gains and Efficiency*

Remarks on the impact of capital gains taxation on economic efficiency are speculative, to say the least. However, two major aspects should be considered: First, the real costs of tax avoidance must be deducted from any gains that can be attributed to preferential taxation. Second, preferential capital gains taxation distorts investment and dividend policy. Such distortions may offset a large part of the gain in market efficiency ascribed to the present system.

The capital gains loophole encourages the investor-taxpayer to buy opportunities for earning income that yield capital gains. For example, investment in certain industries, such as livestock and timber, is favored over investment in retailing or manufacturing. In the former case, the basic product can be sold at a capital gain rather than ordinary income. As a second example, investment is channeled through dummy corporations to permit appreciation of stocks and reduce the proprietor's ordinary income. The result is twofold:

1. Investment is diverted into areas favored by the tax law without regard for the economic profit that the investment yields. Profitability of alternative investments is appraised on the basis of after-tax rate of return; after-tax return depends on legal stipulations as to the eligibility of receipts for capital gains taxation as well as the aggregate of other taxable income. As a result, the valuation of alternative investments depends, not on their real return to the economy, but on a distortion of that return induced by the tax law. Overinvestment will occur in areas in which transactions produce current income at capital gains rates. }

2. Knowledge and execution of transactions in the legal form that



admits of capital gains treatment requires the services of accountants, lawyers and business management. Resources are consumed in tax avoidance. As one tax lawyer has put it, tax benefits can be purchased for a price. In a tax system less open to avoidance a portion of those resources could be mobilized for the production of final goods, increasing the efficiency of the economy.)

How large might such compliance costs be? My estimates are no more than a guess, but if one assumes that private individuals spend ten times what government spends on enforcement of capital gains and that a fifth of tax enforcement effort is associated with preferential taxation of such gains, one must conclude that it is possible that compliance costs for this loophole alone range as high as \$1 billion.<sup>3</sup> While this sum does not look large by comparison with the \$30 billion gap between present production and capacity production that would be possible in a full employment economy, it is clearly an order of magnitude that should not be ignored. Time that accountants, tax services brokers, lawyers, and government workers now spend in coping with tax avoidance made possible by preferential capital gains provisions might be better used in increasing cost accounting efficiency and improving productivity of the economy.

The second aspect of efficiency mentioned relates to the impact of capital gains taxation on corporate dividend policy, corporate investment, and the valuation of corporate securities. How are the key corporate decisions of retention and investment affected by capital gains taxation? The following argument appears to be significant and worth empirical investigation.

Investors in corporate securities recognize that after tax return of securities that are otherwise equivalent will be greater, the greater the proportion of funds that are retained in the company. Earnings that are retained will increase the value of the share. That increase will be taxable at preferred rates in some future period while dividends are taxable today, this year, at ordinary rates. Investors will therefore prefer to supply funds to corporations with high rates of retention rather than corporations with low rates of retention.<sup>4</sup> The cost of capital thus favors companies that retain earnings. These same companies can invest funds in new undertakings that will not meet the test of the market until after a gestation period of several years. By contrast the firm that raises its funds in the market must offer present evidence to the market that funds are to be used to develop products and capacity that will be profitable in the future. The scrutiny

<sup>3</sup> In the fiscal year 1962 IRS expenditures for administration totaled \$450 million.

<sup>4</sup> In an elegant proof that valuation of securities is independent of dividend payout, Miller and Modigliani [21][22] entirely ignore the effect of preferential capital gains taxes on investor dividends.

that new investment proposals receive in the capital market subjects all comers to the same test of economic profitability. By contrast investment of retained funds defers scrutiny to the product market at a later date and permits established firms to invest at less than current market rates. Were both categories of investment demand channeled to the same market, it seems likely that a more efficient allocation of investment funds could be accomplished. To the extent that retained earnings are invested in projects that do not ultimately produce satisfactory yields while funds committed under market scrutiny offer better returns, the present system may lead to unnecessarily high market rates of interest. Investments covered by new issues are required to demonstrate greater profitability than investments covered by retained earnings because the supply of funds in the market has been limited by retention of earnings.

The contention that capital gains are required for liquidity and turnover in the capital market thus ought to be countered by the observation that removal of preferential provisions for capital gains would immediately increase the supply of funds available in the capital market and might potentially create an offsetting number of transactions. (Both assertions require a careful econometric study of the market before they can be validated or rejected.) The increase in supply of funds would be derived from larger dividend payouts that are reinvested in the market by shareholders.

#### *IV. Proposed Changes in the Tax Treatment of Capital Gains*

In the bill now pending before the Senate, little has been done to develop appropriate solutions to the problem of taxing capital gains. The major act of the House was to recognize clearly that present capital gains treatment represents a subterfuge from progressive tax rates for the high-income taxpayer. The House established two categories of assets eligible for preferred tax treatment: preferred assets held more than six months and more preferred assets held for more than two years. Numerous transactions that had been given capital gains treatment in the past were excluded from the more preferred class A assets.

Apparently the authors of the bill feel that some property transactions give rise to income that is more socially desirable than others and which, therefore, should be less heavily taxed. But, there is nowhere in the Committee report a valid justification for that distinction. The new distinction is as arbitrary as the whole capital assets definition and the present tax treatment.

In fact the bill now pending before Congress increases the lock-in effect. It seems likely that investors who might have sold after six

months will hold their assets for another eighteen months to take advantage of the more preferred rate [9]. The shrewd tax dodger who contemplates holding his appreciated assets until death will still be locked-in to his investment. (Admittedly he may be relieved of a psychic burden by the fact that taxes will be applied with a 40 percent rate of inclusion or a maximum of 21 percent tax rather than rates applicable under the existing law, but the fact remains that complete avoidance of tax still looks better than a 21 percent tax.)

(The tax avoidance effects of capital gains taxation can only be eliminated by taxing transactions on capital assets on the same basis as ordinary income and eliminating the possibility of discretionary postponement of tax by some system of accrual taxation or the recognition of interest due on deferred tax liability.) The Congress has moved in the opposite direction on both counts. The new law will increase the rate differential between many capital transactions and ordinary income transactions. The new law will increase the scope for avoiding the tax altogether by granting exemption from the tax to certain elderly persons who sell their homes. To my knowledge no analysis by the Administration or the Congress indicates that the proposed changes in law would provide desirable economic effects in the areas of growth, stabilization, or economic efficiency. Neither can the equity of the proposed changes be defended. If new law goes into effect, the scope for tax avoidance will be enlarged, the differential between actual and nominal tax rates will remain. Taxpayer morale will continue to wilt and the economy will continue to bear an enormous burden of tax compliance costs.

To summarize, preferential capital gains taxation in its present form is destabilizing. Whether full taxation would increase the stabilizing process of the tax system depends on resulting changes in asset price movements. The impact of preferential capital gains taxation on efficiency is determined by a balance between real tax avoidance costs and real economic growth that the tax may facilitate. However, it is unclear that real economic growth induced by the tax is of any appreciable magnitude. The overall impact of taxation on the amount of saving is not known. If there is a strong effect on saving, it is unclear how that effect will be mirrored in new investments. Much of the saving can be absorbed in increasing asset prices. How such changes might ultimately stimulate investment is still in the realm of controversy [19]. And even if increases in investments undertaken because of preferential taxation were to amount to as much as \$1 billion a year, the change in our rate of economic growth would be negligible. The link between preferential capital gains taxation and important quantitative effects on our country's growth thus appears to be one that must be demonstrated, not merely asserted.

## REFERENCES

1. M. J. Bailey, "Capital Gains and Income Taxation" (Brookings Institution, Oct. 29, 1963, in process).
2. W. J. Blum, "A Handy Summary of the Capital Gains Arguments," *Taxes*, April, 1957, pp. 247-66.
3. G. F. Break, "On the Deductibility of Capital Losses under the Income Tax," *J. of Fin.*, 1952, pp. 214-29.
4. ———, "Income Taxes and Incentives to Work," *A.E.R.*, Sept., 1957, pp. 529-49.
5. S. A. Brown, "The Locked-In Problem," *Federal Tax Policy for Economic Growth and Stability*, Joint Committee on the Economic Report, 84th Cong., 1st Sess., Nov., 1955, p. 367.
6. E. F. Denison, *Sources of Economic Growth in the U.S.*, Supplementary Paper No. 13 (C.E.D., 1962).
7. E. S. Domar and R. A. Musgrave, "Proportional Income Taxation and Risk-Taking," *Readings in the Economics of Taxation* (R. D. Irwin, Inc., 1959).
8. G. Gurin, J. Veroff and S. Feld, *Americans View their Mental Health* (New York: Basic Books, 1960).
9. H. H. Hinrichs, "An Empirical Measure of Investors' Responsiveness to Differentials in Capital Gains Tax Rate among Income Groups," *Nat. Tax J.*, Sept., 1963, pp. 224-29.
10. Charles C. Holt and John P. Shelton, "The 'Lock-In' Effect of the Capital Gains Tax," *Nat. Tax J.*, Dec., 1962, pp. 337-52.
11. D. W. Jorgenson, "Capital Theory and Investment Behavior" (mimeograph, 1962).
12. W. F. Lewis, *Fiscal Policy in the Post-War Recessions* (Brookings Institution, 1962).
13. R. L. Richman, "Reconsideration of the Capital Gains Tax—A Comment," *Nat. Tax J.*, Dec., 1961, pp. 402-04.
14. R. E. Slitor, "Problems of Definition under the Capital Gains Tax," *Nat. Tax J.*, Mar., 1957, pp. 26-37.
15. B. W. Sprinkel and K. B. West, "Effects of Capital Gains Tax on Investment Decisions," *J. of Bus.*, April, 1962, pp. 122-34.
16. S. S. Surrey, "Definitional Problems in Capital Gains Taxation," *Tax Revision Compendium*, submitted to the Committee on Ways and Means, 2 (Nov. 16, 1959), pp. 1203-32.
17. R. Likert, *New Patterns of Management* (McGraw-Hill, 1961).
18. B. F. Massell, "Technical Change and the Aggregate Production Function," *Rev. of Econ. and Statis.*, May, 1960, pp. 182-88.
19. T. Mayer, "The Empirical Significance of the Real Balance Effect," *Q.J.E.*, May, 1959, pp. 275-91.
20. J. Meyer and E. Kuh, *The Investment Decision* (Harvard Univ. Press, 1957).
21. M. H. Miller and F. Modigliani, "Dividend Policy Growth and the Valuation of Shares," *J. of Bus.*, Oct., 1961, pp. 411-33.
22. ———, "On Dividend Policy and Market Imperfection," "Comment" by Wm. J. Baumol, *J. of Bus.*, Jan., 1963, pp. 112-15; "Reply," pp. 116-19.
23. R. A. Musgrave, *The Theory of Public Finance* (McGraw-Hill, 1959), pp. 312-36.
24. R. Turvey, "A Tax System with No Company Taxation," *Canadian Tax J.*, Sept.-Oct., 1963, pp. 409-19.

## DISCUSSION

RICHARD A. MUSGRAVE: Theorizing about the incidence of the corporation tax has reached a point where empirical testing is sorely needed. This does not mean simple observation of overall results (e.g., deduce full shifting from the fact that the gross rate of return doubled from the mid-twenties to mid-fifties, or conclude no shifting from the fact that the profit share showed little change) but a serious effort to isolate the effects of tax rate changes on the relevant variables. Such efforts are exceedingly difficult and scarce, and as a recent coauthor of a similar yet quite different study, I welcome Professor Hall into the fold. But, as may be expected, there will be different ideas on how the baby should be raised. While these early efforts are all pretty much in the kindergarten stage, I am left with the feeling that our engine carrying a sign "towards full shifting" seems to me on substantially firmer tracks than Professor Hall's "zero-shifting" express.

In reading Professor Hall's paper, I am aware that space limitations did not permit him to include the degree of detail which undoubtedly will be found in the complete study. However, as it stands, I find it difficult to see how he can sustain his claim that his "results are a strong argument for the priority of the no-shift case, and indeed are largely sufficient as grounds for decision." They strike me as highly inconclusive and, what seems worse, the very nature of the model may be such that no rigorous evaluation is possible.

To begin with the most obvious point, Professor Hall's margin of victory—97.22 percent of deflated hourly output accounted for with the no-shift assumption and 94.55 percent with the full sales shift—seems much too slim to yield more than the most tentative suggestion. Such is the case especially since no statistical measure of significance is given. I hope that such a test can be developed, since a model which does not permit adequate testing of the main result is of modest value. Indeed, throughout the paper very little is said about the quality of the data, their statistical properties, or the nature of the many complex adjustments involved, including especially the physical measure of capital and allowance for changes in the degree of capital use. The data cover a very long and varied period, raising serious doubts as to homogeneity, and the question whether it would not have been better to excluded the depression and war period. The presence of serial autocorrelation in error terms is mentioned, but I do not know how serious this is and how the three models compare in this respect. It is to be hoped that all this will be explained in the full study, that alternative procedures will be explored, and that the implications of adopting one or another type of adjustment will be tested in plain view of the reader. Until then, it remains difficult to evaluate the results.

All this is not to deny that the basic idea of Professor Hall's paper is an attractive one. If the "true" adjustment for technical progress as well as the shape of the production function were known by divine inspiration, obtaining a distinctly superior fit with use of the pre-tax profit data would indeed lend support to the no-shift hypothesis. But with all respect for Professor

Solow's pioneering efforts in this area, his technique of isolating progress and of demonstrating its neutrality is as yet full of unresolved problems. It is, I submit, an exceedingly delicate reed on which to hang an incidence analysis.

The fact that an about equally good fit is obtained with most any production function, including a rather curious linear form, to my mind, does not strengthen the result, but suggests that the apparatus is very crude and insensitive to the phenomenon under investigation. One wonders, for instance, what degree of unneutrality in technical progress would be needed to give a no-shifting result for a Cobb-Douglas function; or, inversely, what adjustment in the production function would give this result with neutral progress.

This much for the no-shifting case. When it comes to the other two cases, I am even more worried. What am I to make of a model which, à la Solow, assumes the competitive market hypothesis underlying the adjustment for technical change and the fitting of the production function, and then proceeds to hypothesize forward or backward shifting of the short-run type, introducing thereby imperfections into the product and/or factor market which are thoroughly incompatible with the general framework of the earlier analysis? The fact that the high explanatory value of the model is barely reduced by this mix of assumptions again underlines the crudeness of the approach.

Moreover, one wonders why only the extreme assumptions of 100 and zero percent shifting are tested. As I see it, there is no reason to assume that Professor Hall's  $R^2$  must decline continuously as the assumed degree of shifting is raised from zero to 100 percent. If the  $R^2$  criterion is accepted (and, as noted before, I doubt the adequacy of this), why not construct the model so as to make the degree of shifting a dependent variable and determine that degree for which  $R^2$  is largest?

In concluding, let me say a word about the alternative model, developed by Professor Krzyzaniak and myself.<sup>1</sup> As distinct from the share approach chosen by Professor Hall, our model focuses on the rate of return. A model is constructed which permits us to predict the rate of return on capital, for the all manufacturing and other groups. Corporation tax rates (unlagged and with a one-year lag) appear as predetermined variables, and the regression coefficient pertaining to them is translated into a measure of shifting. The experiment is repeated for various concepts of capital, with and without adjustment for elements such as inflation and accelerated depreciation. The results throughout show more or less full shifting of the tax; i.e., the gross rate of return rises sufficiently as a result of the tax change to leave the net rate unaffected. The fit of our estimating equation is reasonably good and the tax-rate coefficients are statistically significant in most cases. While we are unable to overcome an unpleasantly high correlation between corporate rate and government expenditure changes, the evidence for a high degree of short-run shifting is nevertheless impressive.

Finally, can the Hall and Krzyzaniak-Musgrave results coexist, or is an agonizing reappraisal needed? As we have noted, a distinction need be drawn between the consequences of a tax change which comes about quickly and

<sup>1</sup> Marian Krzyzaniak and Richard A. Musgrave, *The Shifting of the Corporation Income Tax* (Johns Hopkins Press, 1963).

those which take longer to work out. The former are referred to as "short-run" shifting, even though this type of adjustment may prevail in the long run as well. In particular, if the short-run adjustment succeeds in maintaining the net rate of return on capital at its pre-tax value, there will be no long-run depressing effects on capital formation. It is my understanding that Professor Hall also limits himself to the "quick" type of adjustment. Indeed, his one-year time period seems to be even shorter than ours, since no attempt is made to lag the tax rate. But if it is true that both studies deal with the quick adjustment only, the rate-of-return indicator of our study and the share indicator of Hall's study should give the same result. Such at least is the case if we assume that there can be no significant changes in capital stock during this very short period. The outcomes of the two studies, therefore, are incompatible. At least one of them must be wrong.

At the same time, either result can be reconciled with the observed facts that the gross rate of return on capital doubled, while the profit share showed little change from the late 1920's to 1950's. Since the capital-to-labor ratio increased only moderately during this period and wage rates also about doubled, the Hall result is compatible with the hypothesis of neutral technical change and a Cobb-Douglas production function; our result, according to which the capital share would have fallen without tax, suggests laborsaving technical change with Cobb-Douglas, or neutral change with a somewhat different production relationship.

Such rationalizations are possible, but inconclusive. The observed changes in overall relationships are not a product of tax changes only, but may also reflect other structural changes; e.g., changing bargaining power of unions or control over administered markets. Casual observation without properly isolating the influence of tax change does not reveal the truth. The two models, therefore, must stand (or fall) on their analytical and statistical merits.

JOHN LINTNER: John Brittain's fine paper on "The Tax Structure and Corporate Dividend Policy," in my view at least, clearly represents one of the two most constructive and stimulating contributions to the literature on what these policies are which has appeared in many years.<sup>1</sup> Its focus is on the empirical patterns and regularity of behavior as it can be observed and explained, not on the separate normative issues of what policy "should" be in order best to attain certain specified objectives for the companies or shareholders involved. But this type of behavioral, descriptive, or structural knowledge is of course precisely the input required, in turn, for normative analyses of what public policy with respect to tax rates and tax structure should be in order to accomplish certain goals for the economy or society at large.

Perhaps I am led to this generally favorable reaction to Brittain's work, in part, at least, because it imaginatively builds upon and so nicely complements my own work on corporate dividend policy reported at these meetings some years ago.<sup>2</sup> Whereas I was focusing on the "Distribution of Incomes of

<sup>1</sup> The other contribution I refer to is the discussion of dividend policy (primarily in parts of Chap. 2 and in Chap. 10) by Edwin Kuh, *Capital Stock Growth: A Micro-Econometric Approach* (Amsterdam: North-Holland Pub. Co., 1963).

<sup>2</sup> "Distribution of Incomes of Corporations Among Dividends, Retained Earnings, and Taxes," *A.E.R.*, May, 1956.

Corporations Among Dividends, Retained Earnings and [Corporate] Taxes," he turns to "The [Personal] Tax Structure [Depreciation Allowances] and Corporate Dividend Policy"—and perhaps for clarity the added words should have been included in both our titles. And while I was essentially concerned to establish that dividend policy can best be understood in terms of target payout ratios—which are determined primarily by long-run, equilibrium-type considerations and which once established are relatively stable over substantial periods—he is essentially concerned with establishing the relevance and importance of two factors determining what the target payout will be. Again, in a complementary way, Brittain in the current paper at least is concerned primarily with "aggregative time series structure," while my earlier work was essentially focused on the decision-making process within the firm and introduced the time series of aggregates largely to show that the formulation of dividend policy in terms of progressive partial adaptation to relatively stable long-run targets actually produced much better explanations of aggregative behavior than other formulations which had previously been offered: If "the basic [i.e., simplified] model" of "the dominant determinants of corporate dividend decisions" (*op. cit.*, p. 113, and also p. 107) did so when fitted with strictly invariant targets and speeds of adjustment it would a fortiori do so when the targets were appropriately allowed to shift gradually in response to major changes in any of the principal considerations listed which, on the basis of our field work and studies of individual companies, were reflected in the target payout figures (and correspondingly for the speeds of adjustment which were expected to be relatively more flexible).<sup>3</sup>

One of Brittain's two major conclusions is that the major changes in the levels of personal tax rates (particularly in the higher brackets) which have occurred over the last forty years or so have altered the (weighted-average or aggregate) target payout ratio of American industrial corporations. As a general qualitative conclusion, and also in the context of a macromodel such as he offers, this conclusion strikes me as eminently sound. In particular, I regard it as probably the single most important explanation of the structural shift in a profits-oriented function which he finds between prewar and postwar dividend behavior. In addition to the statistical evidence he offers today, I can add the evidence of our field interviews which indicated that "the importance attached by management to longer term capital gains as compared with current dividend income for its shareholders" (*op. cit.*, p. 104) was one of the major considerations in fixing target payout ratios. More generally, companies were found to seek a "reasonably consistent pattern of action which would both meet the company's needs most of the time and reasonably balance the longer term interests of shareholders in the company and their shorter term interests in current income." (*Op. cit.*, p. 102.)

But while personal tax rates are clearly an important factor in target payout setting for many companies (and in the aggregate), it is clear from interviews that most companies do not adjust their targets except at relatively long intervals, and that several (including some well-known giants) made no

<sup>3</sup> At this time I might also note specifically that the constant term  $a$  was used only as a simple linear term covering certain other features such as premium on growth in dividends, and the greater reluctance to cut than to raise dividends—all of which can (by complicating the rest of the function) be better handled directly in a homogenous form.



change at all in their targets during or after World War II as a result of changing personal tax rates. I consequently regard the "continuous action" on a year-to-year basis which Brittain finds in his tax variable as an aggregation phenomenon, subsuming the different lags (and extents) of the change in target ratios of those companies which did change targets for personal tax reasons, as well as weight shifts among firms and industries, and probably the effects of other excluded variables. In particular, I must part company with him in defining the relevant tax rate in each year as the marginal rate paid on shifting income level which marks off the ninetieth (or any other fixed) percentile of the distribution of dividend receipts—a variable which confounds changes in tax laws with the very different phenomenon of changing income distributions. Again there is interview evidence of the relevance of a company's judgment of income levels of its shareholders in fixing targets, but on an individual company basis these seem to be rather static judgments.

All in all, I must conclude on the personal tax issue that (a) major changes in personal tax rates do affect target ratios for at least a significant fraction of individual companies, and hence for broad aggregates; but that (b) in the context of judging individual company policy on payout targets, Brittain's results probably overstate the extent, and especially the frequency, of the effect; and that (c) the effects of changes in tax laws are again seriously overstated in the context of structural estimates of aggregative behavior since the effects of changing tax rates and (a composite of cyclical, inflationary and trend) changes in income distributions are confounded in his tax term; but that (d) they may nevertheless be entirely acceptable in the context of developing an equation for the specific purpose of the forecasting (as distinct from structural estimation) of aggregates.<sup>4</sup> In addition and very importantly, as previously mentioned, (e) the order-of-magnitude change in personal tax structure during and since the war stands out as probably the major explanation of the structural shift between prewar and postwar dividend policies which is found in aggregate data and in that of individual companies.

Turning now to the other major conclusion of Brittain's paper—that the use of cash flow instead of reported profits provides a better explanation of dividend behavior—I must again register some reservations. In our interviews we had found that "management's confidence in the soundness of earnings figures" (*op. cit.*, p. 104) was an important factor in dividend policy quite generally, and we all know that security analysts and company treasurers alike refer to earnings having better "quality" if depreciation reserves are relatively large, if various adverse contingencies have been conservatively (i.e., generously) charged off or "reserved," and if R and D has been expensed rather than capitalized, etc. At the same time we found companies who had been setting up special reserves by direct charges to surplus to cover inadequate depreciation who insisted this had not affected their adherence to established policies for dividend payments in relation to reported earnings, and others who were getting large amounts of accelerated amortization on facilities clearly

<sup>4</sup>For some of the rationale involved here see, for instance, John Meyer and Robert R. Glauber, *Investment Decisions, Economic Forecasting, and Public Policy* (in press). Also on a related issue see "Is Aggregation Necessarily Bad?" by Yehuda Grunfeld and Zvi Griliches in *Rev. of Econ. and Statis.*, Feb., 1960.

needed and useful for later (post-Korean) operations, who similarly adhered to dividends in line with reported earnings—and statistical analysis of company data confirmed the management's assertions in both types of cases.

Furthermore—although this is not the proper occasion to present my own later research on dividend policies in any detail—there is much cross-section evidence against the “cash-flow” position at least through 1954. As an illustration of the evidence, it seems clear that if cash flow was generally important in company dividend policy, it should show up clearly in cross-sections of such industries as heavy steel, heavy chemicals, special industrial machinery and machine tools, and in the petroleum industry with its large and well-known depletion allowances. But a depreciation variable added to an equation on lagged dividends and reported profits in each of these industries in 1951 and 1954 has a negative sign in seven of the ten cases, and the three positive partial correlations were less than .07. Data were deflated to eliminate size bias, but similar results hold without deflation, and on either basis seventeen other industries show the same general unimportance or perverse effect (reduced adjusted  $R^2$ 's) of depreciation as a separate variable.

The sum of our interview evidence, this type of cross-sectional statistical evidence, and the fact that the profit-lagged dividend model seems to perform very well at least through the mid-1950's (particularly so when allowance is made for a shift in target ratios induced by much higher personal tax rates and/or if profits are adjusted for inventory valuation) lead me to conclude that depreciation allowances per se (or the cash flow) was probably not a true structural variable in dividend policy until rather recently. I am quite prepared, however, to entertain the proposition that cash flows per se have taken on an importance in their own right since “Mr. Smith went to Washington” and facilitated DDB and SYD depreciation, and especially since Mr. Dillon and the Democrats have revamped depreciation lives as the Republicans had not done for whatever reason. It may indeed be that in these recent years managements have been estimating earnings *de facto* from cash flows more or less in line with the theoretical concept of earnings I have been advancing in normative work for several years,<sup>5</sup> and if indeed this should prove on further analysis to have been the case, it would be another instance of corporations, despite undoubted widespread separations of ownership and management, behaving in ways that approached normative allocative optima.<sup>6</sup> But unfortunately, unavoidable delays in getting our data in shape for a direct cross-section test prevent my reporting the results of such tests for recent years or having any firm convictions on this depreciation or cash-flow issue with respect to these recent years. While I am at least open minded on, if not inclined to favor, this view, the case is still unproven—in part because, as Brittain observes, other unincluded variables may be operative. I would particularly nominate risk factors in this connection—including a measure of the sea-change in managements' subjective view of the risk of major depressions—a variable to reflect anything lowering their subjective probability as-

<sup>5</sup> See, for instance, my “Optimal Dividends and Corporate Growth Under Uncertainty,” *Q.J.E.*, Feb., 1964, and references to earlier papers.

<sup>6</sup> “The Financing of Corporations,” Chap. 9 in *The Corporation in Modern Society*, edited by Edward S. Mason (Harvard Univ. Press, 1959).

sessments of its occurrence, including their growing faith in the efficacy of fiscal and monetary stabilization measures and/or their fading collective memory of what a real blizzard would be like. Clearly more work is needed on these issues.

As one final comment, I must confess surprise at Brittain's surprise that he did not find current investment outlays an active factor in determining current dividend payments. I, at least, was not surprised—in part because this result squares with my own conclusions reported earlier and with subsequent further investigations including cross-sections. I consequently regard Brittain's result on this important point as not only sound but as providing important confirmation to the basic rationale of my whole approach in the financial area—which is that target-values for dividend payouts (and debt-equity ratios, liquidity and the relative size of the investment budget) are simultaneously determined in the context of a “comparative dynamic stochastic moving equilibrium,” along with (somewhat more flexible) “speeds of adjustment” to stochastic developments and temporary departures from trend as they may occur. The relative richness of investment opportunities, the character of the competitive and stochastic process within which the firm lives, and so on (cf. *op. cit.*) all affect what the target ratios are, but once set they are not changed lightly or frequently. The evidence is growing that the dividend payment is a top-priority item in the course of shorter term swings, and that companies can and do pay the dividends implied by their relatively stable long-run oriented policies with considerable consistency over substantial periods of time, adapting other phases of their adaptive policies to this near-requirement for dividend outflow over successive short runs, and adapting the long-run dividend target itself only if and as there are major changes in environmental factors and trend-expectations or major adverse changes in the company's fortunes requiring a reorientation of the whole long-run financial strategy of the company.

E. CARY BROWN: While I am sympathetic with the general thrust of Professor Martin David's paper—that the economic effects of full capital gains taxation need not trouble us unduly—I find it necessary to place a number of question marks at the end of some of his conclusions and analyses. Surely, we must be careful in appraising the economic effects of such a change, since policy formulation necessitates our weighing these changes against those arising on distributional grounds.

In reading over his paper I had some difficulty in ascertaining just what revenue structures are being compared. While Professor David is certainly well aware of this ambiguity, and utters many caveats, it is still unclear what are the two alternative systems of taxation with equal revenue yields. It is one thing to assume a major shift in tax policy that would not change economic behavior; it is another to compare revenue consequences when economic behavior changes markedly, as would be likely under the two cases dealt with here. Were there to be no distinction in taxation between capital gains and dividend income, corporations would have less pressure for retention of earnings, dividends would presumably be higher, and personal tax revenues would

correspondingly swell. Therefore, general rate changes could be less than they otherwise would be, were corporate dividend behavior to remain unaffected. This difference will have an important, if not crucial, bearing in appraising economic effects.

Basically a shift in tax structure of the kind analyzed by Professor David involves some increase in the taxation of property income in favor of income from personal effort, with, of course, redistribution of tax among various forms of property income. In considering the consequences for economic growth two issues arise—the effect on aggregate saving and on the quality of investment. I am not sure whether or not Professor David believes that this shift will change aggregate private saving in the community. In principle the consequences are unclear. There will be an increase in aggregate disposable income and it seems likely, or possible, that total private saving could actually decrease, given the fact that substitution effects would work against saving. One important question, then, as to economic growth is the effect of this tax change on aggregate saving, since that will determine, basically, the amount of resources available for investment at full employment. Whether or not there would be sufficient full employment investment in any event would depend primarily on other economic policies, although the probable reduction in corporate saving may make this problem more difficult.

If this were, in fact, the case, the smaller saving (higher consumption) would result in a smaller potential stock of capital at any point in time and lessen our production possibilities. The degree to which they are lessened is unsettled, since there is no reasonably clear-cut answer to the weights to be attached to the basic determinants of economic growth. I am surprised, though, to find Professor David's references to the effect of the capital stock on growth rates to be those that emphasize insignificant results, in view of the possibility raised by Solow and others that the embodiment of technologic change in capital formation can make a significant difference in rates of growth. We should not yet be prepared, therefore, to accept so easily the apparent slight consequences of heavier capital gains taxation on potential economic output. It may be as small as Professor David says, but it also may be considerably larger.

As to the consequences on economic stability of heavier capital gains taxation, I would agree with Professor David that it probably makes relatively little difference. As long as the realization concept is followed—and I see little hope yet for any other technique—the orders of magnitude cyclically between the two tax structures are relatively small, and certainly not decisive.

The efficiency case against the present treatment of capital gains, however, seems to me to be a particularly powerful one. There are the important and obvious inefficiencies that arise from arbitrary classifications of income as capital gains that Professor David properly points out to have serious dead-weight losses. There is also the distortion from the present method in portfolio selection toward capital-gain-yielding investments rather than those yielding ordinary income. This distortion may be an appropriate offset to another distortion arising from our inability to overcome private risk. But this assumes that assets on which there are potential capital gains are also the

most risky. Do we have evidence to suggest much support for this position? On what basis can we assert that a corporation paying out a small proportion of its income as dividends is necessarily riskier than one paying a high proportion? Or than a similar venture operating as an unincorporated enterprise? We must also remember that a very large fraction of capital gains are realized in real estate transactions—an area of risk, to be sure, but not necessarily of the same degree of risk as a vast number of other ventures. Indeed, some become almost riskless through leasing arrangements. Until there is some linkage established—and it might be a fruitful area in which work could be performed—the effects of higher taxation on the realization or generation of capital gains may have little to do with risk-taking.

Even were this connection established, Professor David properly points out that the higher taxation of income from risky assets may not adversely affect risk-taking, provided we press for symmetry in treatment of losses from them.

Professor David is also concerned about the distortions of transactions created by the concept of realization and urges the need for techniques of meeting them through some method of accrual taxation. Other, less extreme devices are also available, based on the treatment of gifts or bequests as a realization. There is still a simpler device which I would like to propose as one method of meeting this problem: to include in income an imputed interest at a rate of, say, 4 percent, on the adjusted basis of the asset. Such a method would be easy to operate and would tend to spread income from capital gains more evenly over the years without as much tax postponement as arises under a realization concept. Thus, given some initial basis, a taxpayer would include at least 4 percent in income. If received as dividends, the basis would remain unchanged; if no dividends were received, the amount included in income would be added to the adjusted basis of the asset. End-year valuations would be avoided. All that would be necessary would be some starting point, which all assets must have.

The last point I should like to make has to do with Professor David's view that the present treatment of capital gains favors existing corporations in their sources of funds and discriminates against funds raised through the capital market. It is clear, as has been said before, that present methods put more pressure for retention than would the alternative method he discusses. Yet, once an existing corporation has made the decision to retain a dollar of earnings, it can offer its stockholders no more than a new company or a company with a new flotation can. They can each hold out potential capital gains to their stockholders on basically an equal footing. I do not see this as an additional distorting effect resulting in a higher cost of capital to the equity seller.

In sum, I am sympathetic with the concern Professor David expresses for the distortions produced by our present method of treating capital gains. I am not as sanguine about its effects on economic growth as is he. Since I believe that there would be an equity improvement from full capital gains taxation, I would be happier were his economic analysis to allay all my fears.

# FINANCIAL FACTORS IN BUSINESS CYCLES

## FINANCIAL MARKETS IN BUSINESS CYCLES: A SIMULATION STUDY

By FRANK DE LEEUW

*Board of Governors of the Federal Reserve System*

This paper is a sequel to an earlier study of postwar United States financial behavior. The earlier study presented a nineteen-equation quarterly model of supplies and demands in major financial markets. It included detailed discussion of individual supply and demand relationships; but it did not deal with the properties of the model as a whole. The present paper explores the properties of the complete model by means of a series of simulation runs.

Some of the questions for which the simulation runs give the model's answers are: What are the effects on interest rates and financial portfolios of a \$500 million open market purchase or sale by the Federal Reserve System, after allowing for all simultaneous interactions within financial markets? How are these effects distributed over time? Do they vary depending on whether the economy is growing steadily or is in a recession? How does a recession followed by a recovery affect interest rates and financial portfolios? What are the effects of changes in reserve requirements and in the discount rate?

The two major building blocks on which the paper rests are the underlying model of financial behavior just mentioned and a computer program to simulate economic models. The computer program, written by Ann Walka at the Federal Reserve Board (who also supervised the actual simulation runs), is a general program for simulating the behavior of variables related through a set of simultaneous equations, with provision for long and complex lags and certain kinds of non-linearity. A description of the program is available on request to Mrs. Walka.<sup>1</sup>

The paper is in three parts following this introduction. The first part is a brief and necessarily incomplete summary of the model. The second is the core of the paper, a description of the simulation runs and an attempt to account for the principal results in terms of the relevant properties of the underlying equations. The final part contains a brief summary and some remarks on the uses of the simulation results.

<sup>1</sup> A description of the model will appear in a volume describing the S.S.R.C.-sponsored quarterly model of the United States economy, of which it is a part. Copies of the description are available from the author. The paper also owes much to the support and patience of the Board's Division of Data Processing. The author, of course, takes sole responsibility for the views expressed in the paper.

Two caveats seem important enough to record before proceeding to the body of the paper. The first is that the paper does not deal with the effects of financial market developments on GNP and its components. It is an investigation of financial market behavior for given paths of GNP and monetary policies. In this respect it follows the usual procedure for investigating a single market or sector of the economy.

The second caveat, to which the concluding section of this paper will return, is that the simulation results below are not predictions or even very confident guesses about how financial markets actually behave. As the description of the underlying model makes clear, the equations of the model are a set of tentative empirical explorations rather than a confirmation or test of any well-developed theory of financial behavior. The simulations below draw out the implications of the equations of the model as a group in a way that may suggest weaknesses, contrasts with other studies, and directions for further work. Hopefully the results have some relation to the actual behavior of credit markets; but it is much too early to say how close the relation is.

### *I. The Model*

The model deals with supplies and demands for a group of seven inter-related financial "commodities": bank reserves, currency, demand deposits, time deposits, U.S. securities, and two broad aggregates labeled "savings and insurance claims" and "private loans and securities." Supply and demand equations for these "commodities" determine their "quantities," or dollar amounts, and their "prices," or interest rates. These dollar amounts and interest rates are the endogenous variables of the model.

The exogenous variables of the model fall into two groups. First of all, there are policy variables. The sum of unborrowed reserves plus currency—that is, the major central bank liabilities minus loans to member banks—is an exogenous "open market" variable; and the supply equation for the reserve and currency markets consists, instead of a behavioral relationship, of an identity defining unborrowed reserves plus currency as equal to this exogenous magnitude. Required reserves ratios are also exogenous, entering the demand side of the reserve market through another identity relating required reserves to deposits multiplied by these exogenous ratios. The discount rate, the ceiling rate on time deposits, and changes in the maturity structure of the federal debt are other policy variables which enter into various behavioral equations of the model.

The second group of exogenous variables consists of GNP, some of its major components both on the income and expenditure sides, and the federal debt. These are endogenous in the complete S.S.R.C. model of which the financial model is a part; and they were treated as endog-

enous in the two-stage least squares estimation of the financial equations. The simulations below, as already noted, abstract from the effects of credit market developments on incomes and trace out the paths of financial stocks and interest rates for given paths of GNP and its components.

The identities relating required reserves to deposits and defining unborrowed reserves plus currency as exogenous are two of the central equations of the model, imposing limits on some weighted sum of the public's deposit and currency holdings and thereby affecting these and other dollar amounts and levels of interest rates. Most of the other equations are behavioral relationships dealing with quarterly changes in a sector's holdings of an asset or liability. A "typical" form for one of these equations is the following, in which  $A$  is a sector's holdings of an asset or liability,  $W$  is a measure of (or proxy for) a sector's wealth,  $r$ 's are interest rates, and  $F$  is some short-run constraint on a sector's spending (for example, quarterly gross income for the public or the net inflow of savings for nonbank financial institutions):

$$\frac{\Delta A_t}{W_{t-1}} = a + b_1 \frac{A_{t-1}}{W_{t-1}} + b_2 r_t + b_3 r'_t + \cdots + b_i \frac{F_t}{W_{t-1}}$$

Changes in holdings of an asset, according to this equation, depend on the lagged level of holdings, interest rates, and short-run constraints on spending. The dollar variables in the equation appear not as absolute amounts but as ratios to "wealth." The public's holdings of currency, demand deposits, and time deposits, bank holdings of excess reserves, and bank borrowing from the Federal Reserve System are some of the equations following this form.

Of the exceptions to this form of behavioral equation, the most important are three equations dealing with relationships among interest rates. One of these relates changes in the average rate paid on commercial bank time deposits to the lagged rate, other market rates, the ceiling rate, and the ratio of bank loans and private securities to deposits. A second equation, roughly similar in form to the first, deals with changes in the rate on private loans and securities. A third equation relates the spread between the long-term U.S. rate and the Treasury bill rate to current and past long-term rates and changes in the maturity composition of the federal debt.

As this description makes clear, the solution of the model for the nineteen endogenous variables in any one quarter depends not only on current exogenous variables but also on a large number of lagged endogenous and exogenous variables. Consequently, the simulations below each begin with a set of assumed initial conditions. These initial conditions plus assumed values for exogenous variables in the first quarter



make possible a first-quarter solution for the nineteen endogenous variables. The elements of this first-quarter solution then become part of the set of initial conditions for the second quarter; and these new conditions together with exogenous variables for the second quarter make possible a second-quarter solution. The simulations thus trace out the estimated paths of endogenous variables corresponding to the assumed path of exogenous variables, given a set of initial conditions.

## II. *The Simulations*

*Equilibrium-growth.* The first simulation run began with an arbitrary set of initial conditions; and then (1) let each exogenous dollar variable (GNP and its components, unborrowed reserves plus currency, etc.) move up at 1 percent per quarter<sup>2</sup> and (2) let each other exogenous variable (required reserve ratios, the discount rate, etc.) remain constant. The purpose of the simulation was to see if the estimated time-paths of the endogenous variables reach a steady-growth equilibrium and, if they did, to use the equilibrium values as a base against which to measure the effects of changes in a single exogenous variable. The "arbitrary" set of initial conditions actually consisted of first-quarter 1955 values for most variables and values closer to current ones for the rest (for example, the ceiling rate on time deposits was set at 4 percent rather than the  $2\frac{1}{2}$  percent of 1955).

After several years of simulation, the time-paths of the endogenous variables approached very close to constancy for interest rates and steady growth for dollar amounts. "Equilibrium" values of key endogenous variables appear in Table 1. These values are based on a GNP level which rises from \$465 to \$500 million during the simulation quarters, required reserve ratios of .15 against demand deposits and .05 against time deposits, a discount rate of  $3\frac{1}{2}$  percent, a ceiling rate on commercial bank time deposits of 4 percent, and similarly familiar values for other exogenous variables.

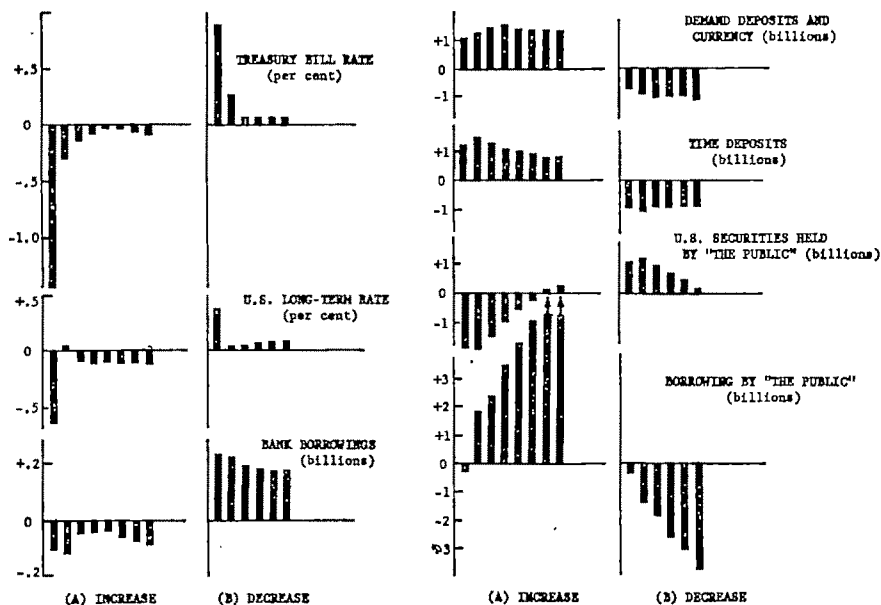
Examination of the table reveals that interest rates are not completely constant, nor do all dollar variables grow by 1 percent per quarter. The greatest departures from these conditions are for those financial stocks which did not fit into a "stock-adjustment" framework in the underlying model; business and household borrowing (shown in the table) and savings and insurance claims (not shown) all of which rise by appreciably more than 1 percent per quarter. Bank holdings of U.S. securities (not shown), closely related to nonfinancial borrowing, fall steadily during the equilibrium quarters. And a slight fluctuation is detectable in bank borrowing from the Federal Reserve and in the Treasury bill rate.

<sup>2</sup> One exception is the federal debt, which is moved up at only half a percent per quarter. A 1 percent per quarter increase would imply a federal deficit of \$8 to \$10 billion each year.

TABLE 1  
VALUES OF SELECTED FINANCIAL VARIABLES, EQUILIBRIUM-GROWTH SIMULATION

	QUARTERS								
	0	1	2	3	4	5	6	7	8
1. Interest rates (percent)									
Treasury bill rate.....	3.14	3.17	3.19	3.19	3.17	3.15	3.11	3.08	3.04
Long-term U.S. rate.....	4.25	4.28	4.30	4.32	4.33	4.34	4.34	4.34	4.34
2. Commercial bank balance-sheet items (billions of dollars)									
Total reserves.....	17.88	18.09	18.31	18.51	18.73	18.94	19.15	19.37	19.57
Borrowed reserves.....	.09	.10	.11	.12	.12	.13	.13	.13	.12
3. Nonfinancial public balance-sheet items (billions of dollars)									
Demand deposits and currency..	137.6	139.1	140.6	142.1	143.5	145.0	146.4	147.8	149.1
Time deposits at commercial banks	75.2	76.6	77.9	79.2	80.6	82.0	83.5	85.0	86.6
U.S. securities.....	102.6	104.3	106.1	107.9	109.7	111.5	113.3	114.9	116.5
Debt to financial institutions....	256.2	263.3	270.8	278.6	287.0	295.6	304.7	314.2	324.1

CHART I  
EFFECTS ON FINANCIAL VARIABLES OF A \$500 MILLION CHANGE IN  
"UNBORROWED RESERVES PLUS CURRENCY"



The equilibrium quarters seem close enough to constancy of interest rates and 1 per cent per quarter growth for financial stocks to suggest that at least for some initial conditions the model approaches a stable rather than a continuously oscillating or explosive path. At the same time, they seem free enough from major disturbances to serve as a base against which to measure the effects of exogenous "shocks."

*A Change in "Unborrowed Reserves Plus Currency."* The next two simulation runs begin with the initial conditions of the equilibrium run and introduce (1) a \$500 million increase and (2) a \$500 million decrease in the exogenous open market variable of the model, unborrowed reserves plus currency. In the first simulation, unborrowed reserves plus currency are increased by \$500 million more than 1 percent between quarter 0 and quarter 1 of the equilibrium run and by 1 percent thereafter; in the second simulation, they are increased by \$500 million less than 1 percent between quarter 0 and quarter 1 and by 1 percent thereafter. All the other exogenous variables are the same as in the equilibrium run. The differences between time-paths in these simulations and time-paths in the equilibrium run appear, for major endogenous variables, in Chart I.

Two results which stand out in the chart are (1) the large but temporary effect of the open market moves on interest rates; and (2) the immediate and lasting effect of the open market moves on the money supply and time deposits. Many of the other results can be accounted for in terms of these results. Thus, movements in the public's holdings of U.S. securities are probably a stock-adjustment response to interest rate changes; and movements in borrowing by the public are mainly a response to interest rate movements and movements in "desired" levels of deposits. There is no stock-adjustment mechanism to slow down the rate of change of the public's borrowing (the stock-adjustment framework did not work for this relationship), with the result that the open market move affects the flow by a roughly constant amount and the stock diverges continually from equilibrium values. Changes in bank borrowing are influenced by interest rates (positively) but also by the public's deposit holdings (negatively) and by the public's borrowing (positively); the net outcome in these simulations is a movement to a new level followed by some fluctuation.

The major results relating to interest rates, money, and time deposits are those which would occur in a very simple model of financial behavior. Suppose the supply of money is exogenous and the demand for money is a stock-adjustment relationship of the form

$$\Delta M_t = \alpha - b_1 M_{t-1} - b_2 r_t$$

in which  $M$  is the money supply and  $r$  "the" interest rate. Solving for  $r_t$  gives

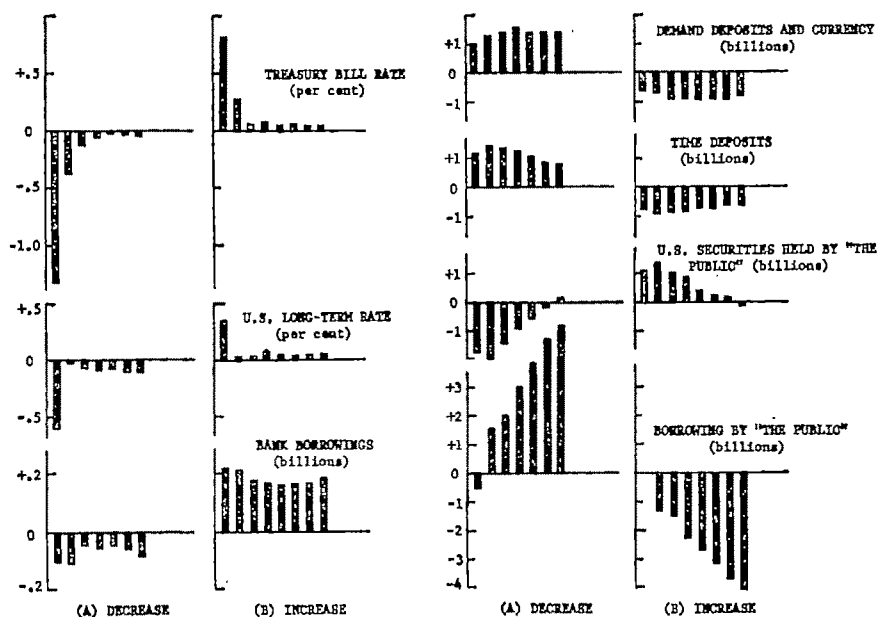
$$r_t = \frac{\alpha}{b_2} - \frac{1}{b_2} [\Delta M_t + b_1 M_{t-1}]$$

An abrupt increase in  $M$  to a new level is accompanied, according to this equation, by an initial change in  $r$  equal to  $-\Delta M/b_2$  and a permanent change equal to  $b_1$  times the initial change. If  $b_1$ , the "speed of adjustment" of the stock of money is of the order of .10 or .15, as it is in the money-demand equations of the model, the initial impact on  $r$  of a change in  $M$  will be seven to ten times the final impact.

The major results of these simulation runs are thus results which two features of the underlying model might lead us to expect. The first of these is the presence of an exogenous supply variable; the second is the stock-adjustment form of many of the demand equations.

The "money multiplier," or ratio of the final change in demand deposits plus currency to the original \$500 million open market change, is between 2 and 3. The required reserve ratio applicable to demand deposits of member banks is .15 in these simulations, and that to currency, of course, is one. With the change in deposits about  $4\frac{1}{2}$  times the

CHART II  
EFFECTS ON FINANCIAL VARIABLES OF A HALF PERCENT CHANGE IN  
RESERVE RATIO REQUIRED AGAINST DEMAND DEPOSITS



change in currency and with member bank deposits about five-sixths of total deposits the required ratio of central bank liabilities to money is about .28 and we might expect the multiplier to be the reciprocal of .28, or about  $3\frac{1}{2}$ . It is lower than  $3\frac{1}{2}$  for two reasons: the cushioning effect (at least in these simulations) of member bank borrowings and excess reserves, and the reserves required against time deposits.

*A Change in Required Reserve Ratios.* The next two simulations trace out the effects of a change in the required reserve ratio against demand deposits from .15 to (a) .145 and (b) .155 in quarter 1 of the equilibrium run. The new reserve ratios continue to apply in subsequent quarters. All other exogenous variables, including unborrowed reserves plus currency, are kept at their equilibrium run levels. Chart II illustrates the time-paths of key endogenous variables.

The results are strikingly similar to those of open market changes. A series of recent papers have addressed themselves to the question of the differential impact of open market operations and reserve requirement changes,<sup>3</sup> especially to the impact on bank asset composition. The simulations are consistent with one of the major hypotheses in these papers; namely, that reserve ratio increases as compared to equivalent open

<sup>3</sup> Joseph Aschheim, "Restrictive Open-Market Operations Versus Reserve-requirement Increases: A Reformulation," *Econ. J.*, June, 1963, pp. 254-66; and the series of earlier papers referred to in that article.

market sales involve a relative shift away from U.S. securities on the part of banks (Chart II does not illustrate the shift). The estimated shift is quite small, however, and according to the simulations the aggregate results of the two kinds of policy are the same (the distributional effects among banks are outside the scope of the simulations).

In both sets of simulations, incidentally, the initial effects of a policy of tightening credit are smaller than those of a policy of easing credit. The reason lies in the behavior of bank borrowing; borrowing has a lower limit of zero but has no rigid upper limit.<sup>4</sup> It can therefore cushion the effects of a credit shortage to a greater extent than it can cushion the effects of an easing of credit. To put it another way, once bank borrowing falls to zero further increases in unborrowed reserves cannot be used to repay bank indebtedness to the Federal Reserve System. The alternative uses which banks must find for these reserves have a greater effect on interest rates than does repayment of debt.

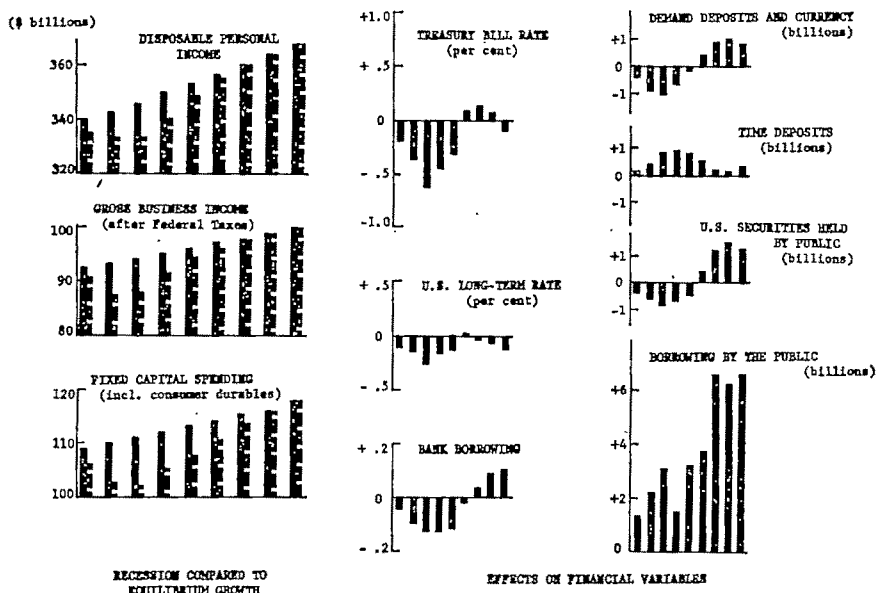
*A Recession and Recovery.* The next simulation traces out the effects on financial markets of a "typical" recession and recovery in GNP—abstracting, as in all the simulations, from the effects of credit market changes on GNP and its components. The recession begins in quarter 1 of the equilibrium run, reaches a trough in quarter 2, and is back on the steady-growth path by quarter 8. The recession affects GNP, personal and business income, and capital spending. Values of monetary policy variables are unchanged from their equilibrium-run values in the simulation. Chart III depicts the recession recovery, and illustrates its consequences by showing differences between the time-paths of endogenous variables in recession recovery and time-paths in equilibrium growth.

Three results of the simulation seem to stand out: (1) the drop in interest rates during the recession, with turning-points lagging slightly the turning point in income; (2) the drop in demand deposits plus currency largely offset by a rise in time deposits during the recession, again with lagged turning points; and (3) the secondary fluctuations in many of the financial variables—GNP and its components have only one turning point but interest rates and some financial stocks have two during the simulation period.

The first of these results is consistent with a model of financial behavior only slightly more complex than the equation discussed above. In that equation the change in money holdings (with money exogenous) depended on the lagged stock and on "the" interest rate. We need now to add current and lagged income to the determinants of changes in

<sup>4</sup> In the simulations the lower limit was imposed by replacing the bank borrowing equation by the equation "borrowing=0" whenever the solution indicated negative borrowing. Since a lagged value of zero for borrowing was not possible in the simulations, an additional rule was followed; namely, using \$100 million for lagged borrowing wherever the actual lagged value was \$100 million or less.

CHART III  
THE EFFECTS ON FINANCIAL VARIABLES OF A RECESSION AND RECOVERY



money holdings (both with a positive influence) and solve for the interest rate. Following the form of a number of equations in the model, let  $Y_t^*$  be some weighted average of past incomes and  $Y_t$  be current income. An equation of the form

$$\frac{\Delta M_t}{Y_t^*} = a - b_1 \frac{M_{t-1}}{Y_t^*} - b_2 r_t + b_3 \frac{Y_t}{Y_t^*}$$

leads to

$$r_t = \frac{a}{b_2} - \frac{1}{b_2 Y_t^*} (\Delta M_t + b_1 M_{t-1}) + \frac{b_3}{b_2} \frac{Y_t}{Y_t^*}$$

$r_t$  will vary positively with both the cyclical position of income  $Y_t/Y_t^*$  and with the general level of recent income in this relationship. Thus three features of the nineteen-equation model underlying the simulation—namely, an exogenous supply variable, a stock-adjustment framework, and a demand equation depending on current and lagged income as well as interest rates—seem sufficient to account for one of the main results of the recession-recovery simulation.

One further extension of the equation just discussed—and again, one which corresponds to a feature of the underlying model—is sufficient to account for the opposite and lagged movements in demand and time deposits. Suppose the exogenously supplied  $M$  is split into two endog-

enous components, one relatively sensitive to current and lagged incomes and the other relatively sensitive to interest rates.<sup>5</sup> The sum of the demand equations for the components will look like the one just presented. As current and lagged income and interest rates fall demand will shift to the interest-sensitive component; as they rise demand will shift back.<sup>6</sup>

The third major result of the recession-recovery simulation, the secondary fluctuations in many of the financial variables, cannot be accounted for quite so simply. One possibility is that they are due to the interrelationships among the various interest rates in the model. To understand how these interrelationships might give rise to fluctuation, suppose money holdings depend positively on income and negatively on two interest rates,  $r$  and  $r'$ .

$$\frac{\Delta M_t}{Y_t^*} = a_1 - b_1 \frac{M_{t-1}}{Y_t^*} - b_2 r_t - b_3 r'_t + b_4 \frac{Y_t}{Y_t^*}$$

Let us suppose further that one of the interest rates depends among other things on lagged values of the other. A very simple equation of this kind is

$$r'_t = a_2 + c_1 r_{t-1} + c_2 r_{t-2}$$

A "shock" in  $Y$  or  $M$  will lead to fluctuations in  $r$  in this model provided that

$$\frac{b_3}{b_2} (c_1)^2 < 4c_2$$

If  $b_3$  and  $b_2$  are of the same order of magnitude and if  $c_1$  is less than one and not too much greater than  $c_2$ , then a fluctuating solution is very likely.

In the equations of the model, yields on private securities and time deposits both follow U.S. security yields (among other variables) with a distributed lag; and demands for deposits depend in part on these lagging yields. The secondary fluctuations in the recession-recovery simulation could well be due to these features of the model.

Without further specification, however, these features do not explain the fact that secondary fluctuations appear strong in the recession-recovery simulation and negligible in earlier simulations. This paper will

<sup>5</sup> Or alternatively, suppose that  $M$  is endogenous, some other asset  $M'$  is also endogenous, and that

$$R_t = k_t M_t + k'_t M'_t$$

where  $R$ ,  $k$ , and  $k'$  are exogenous.

<sup>6</sup> Incidentally, the fact that both time deposits and demand deposits plus currency exceed their equilibrium values at the end of the simulation period is due to a shift out of currency and a rise in bank borrowing.



not attempt an explanation of the finding, but will simply offer the conjecture that it may have to do with the fact that the recession "shock" is temporary, since income eventually recovers its old trend, while the open market "shock" permanently shifts the trend of unborrowed reserves plus currency.

*Open Market Operation in Recession.* The question of whether an increase in unborrowed reserves plus currency has a different impact during a recession than during a period of normal growth is one for which Chart IV gives some simulation results. According to these results, the initial impact on interest rates is somewhat greater during a recession, the impact on bank borrowing is somewhat smaller, while the impact on currency, demand deposits, and time deposits is about the same.

The nonnegative restriction on borrowing probably accounts for these results. The earlier discussion of reserve ratio changes described above discussed the effect of this lower limit on bank portfolio choices; here it is only necessary to add that banks are initially closer to zero indebtedness in the recession-recovery simulation than in the equilibrium-run, so that banks have less debt to repay before reaching zero.

*Discount Rate Changes.* The final three simulation results to be presented have to do with the effect of changes in the discount rate. The first two trace the effects of raising and of lowering the rate by half a percent during equilibrium-growth; the third with the effects of lowering the rate during a recession. The results appear in Chart V.

The first of the three simulations, in which the discount rate is raised by half a percent, causes bank borrowing to fall rapidly to its lower limit of zero and produces moderate changes in other endogenous time-paths. Interest rates rise temporarily and then fall back nearly to old levels, and deposits fall along with the fall in bank borrowing.

The next two simulations, in which the discount rate is lowered in equilibrium and in a recession, produce highly unexpected results. Bank borrowing rises continually during the period shown in the chart and currency and demand deposits rise as well. An extension of one of the simulation runs (the middle one) beyond the period shown in the chart reveals that bank borrowing eventually reaches a peak and begins to decline, but not until three years (and \$700 million) after the change in the discount rate.

The explanation of this result in terms of the model is rather complex. According to the bank borrowing relationship in the model, banks' "reluctance to borrow"—i.e., the negative influence of the lagged stock of borrowing—is much smaller when the flow of borrowing by the public is heavy (in relation to deposit inflow) than when it is light; and in the basic equilibrium-run simulation, borrowing by the public is relatively

CHART IV  
THE EFFECTS ON FINANCIAL VARIABLES OF A \$500 MILLION INCREASE IN  
"UNBORROWED RESERVES PLUS CURRENCY"

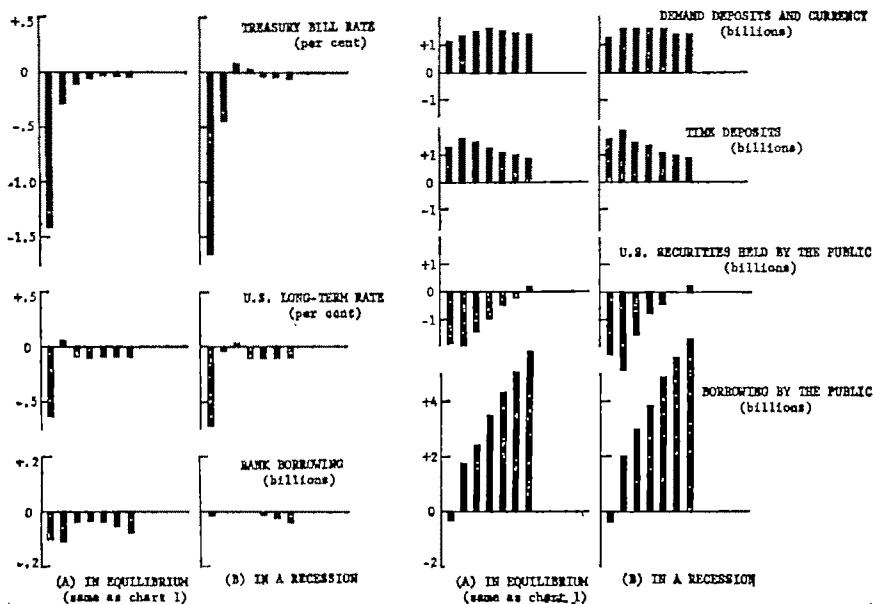
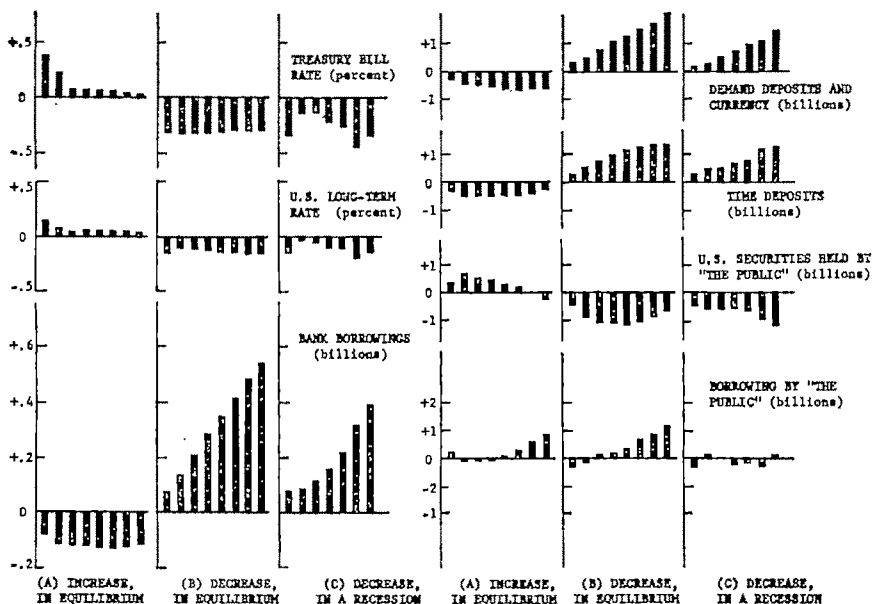


CHART V  
EFFECTS ON FINANCIAL VARIABLES OF A HALF PERCENT CHANGE IN THE DISCOUNT RATE



heavy. Given this underlying flow of borrowing by the public, the effect of the drop in the discount rate is very large; if borrowing by the public were smaller, the effect of a drop in the discount rate would be far smaller.

Whatever the explanation of these results, they are doubtless far different from what most monetary experts would expect. One of the conclusions of the simulation runs, then, is that, according to the model, changes in the discount rate have surprisingly large effects on credit markets. The departure from a priori expectations certainly seems great enough to warrant a re-examination of the model's relationship for bank borrowing and its rather unsatisfactory treatment of borrowing by the public.

*Other Simulation Runs.* Although the simulation runs so far described are probably sufficient to draw out many of the implications of the underlying model for financial market behavior, they far from exhaust the possibilities. Experiments with the effects of changes in the federal debt and its maturity and of changes in the maximum rate on time deposits; experiments involving monetary "strategies" rather than single shocks; and experiments involving random error terms rather than exact relationships would disclose other results of interest. Some simulations along these lines have been run, but space limitations prevent any discussion of their outcomes.

### III. *Concluding Remarks*

The simulations which have been discussed suggest a number of propositions about financial markets. The principal ones relate to: (a) the large and immediate effects of open market operations and reserve requirement changes on interest rates and deposits; (b) the transience of much of the interest rate effects of these changes; (c) the small size of the "monetary multiplier"—that is, of the change in demand deposits and currency divided by the open market change; (d) the lagged response of interest rates and financial stocks to a recession; (e) the secondary fluctuations in financial markets induced by a recession and recovery; and (f) the strong effects of reductions in the discount rate. As the introduction to this paper noted, these propositions are not intended as realistic predictions of how financial markets behave. They are a set of hypotheses about financial market behavior suggested by the underlying, exploratory model and abstracting from the effects of financial markets on markets for "real" goods and services.

Although these results are not intended as predictions, they are intended to be one step in improving our understanding of financial behavior. It seems fitting, therefore, to close this paper by indicating some further steps that might be useful to take.

The first of them is to draw out the implications of other existing models of financial behavior and compare them with the simulation results of the present study and with each other. From these comparisons would emerge some impression of the general consensus of current empirical work with respect to such propositions as the ones just listed.

Testing the present model and other models against developments subsequent to the period in which they were fit is a second and an indispensable step in judging and improving their accuracy and consequently in appraising the propositions about market behavior which they imply.

At the same time, it would probably be quite useful to try to devise methods of testing directly some of the propositions emerging from the present study and other models. It might be possible, for example, to pick out quarters of large open market purchases or sales, allow roughly for other effects on interest rates in those quarters, and obtain a direct estimate of the immediate impact open market operations have on interest rates. Confirmation or rejection by such direct testing of some of the propositions deduced from systems of equations might greatly accelerate the process of eliminating spurious results, correcting for major biases, and developing a serviceable model of financial markets.

# LONGER WAVES IN FINANCIAL RELATIONS: FINANCIAL FACTORS IN THE MORE SEVERE DEPRESSIONS

By HYMAN P. MINSKY  
*University of California, Berkeley*

## I. Introduction

In his 1959 testimony before the Joint Economic Committee Professor Abramowitz stated that "it is not yet known whether they (the long swings) are the result of some stable mechanism inherent in the structure of the U. S. economy, or whether they are set in motion by the episodic occurrence of wars, financial panics or unsystematic disturbances."<sup>1</sup> Quite casually financial panics were identified as exogenous, episodic rather than as endogenous, systemic events. The argument in this paper is that the stable mechanism which has generated the long swings centers around the cumulative changes in financial variables that take place over the long-swing expansions and contractions.

The long swings reach a climax in a deep depression business cycle. Friedman and Schwartz's chronology of deep depression cycles coincides with Abramowitz's chronology of severe contractions that occur at the end of the long waves. Although the evidence for the monetary explanation of mild depression cycles is admitted to be tenuous by Friedman and Schwartz,<sup>2</sup> the evidence that money is a significant part of the mechanism generating a deep depression is strong. The long-wave hypothesis cannot be distinguished from the hypothesis that there are two types of business cycles, mild and deep depression, and that the conditions which must be satisfied if a deep depression cycle is to take place are generated over a number of mild depression cycles.

During the expansion phase of a long swing, or alternatively over a period in which only mild recessions occur, systematic changes in the financial structure occur. These reflect the financial aspects of high

<sup>1</sup> United States 86th Cong. Joint Economic Committee, *Employment and Growth and Price Levels, Hearings Part 2 Historical and Comparative Rates of Production, Productivity and Prices*, p. 12.

<sup>2</sup> M. Friedman and S. J. Schwartz, "Money and Business Cycles," *Rev. of Econ. and Statis.* Sup., Feb., 1963, p. 55. "The case for a monetary explanation is not nearly so strong for the minor U.S. economic fluctuations. . . ." Friedman and Schwartz make a case for a monetary explanation of mild depression cycles by arguing that: "Is not a common explanation for both more appealing than separate explanations, especially when there is no well-tested alternative separate explanation?" (P. 55.)

rates of growth by leading sectors, revaluation of assets to allow for growth expectations and the growing financial layering at the same time as ultimate liquidity<sup>3</sup> grows slowly if at all.

The exact nature of the changes that take place in financial variables during long-swing expansions and the significance of these changes are sensitive to institutional arrangements. Financial institutions and usages evolve, both in response to market forces and as a result of administrative processes and legislation. Thus the exact course of financial variables over the phase of the longer waves differs greatly among the various observed cycles. This makes it quite useless to engage in sophisticated statistical analysis, if the object is to determine the significance of financial factors in the cyclical behavior of the economy, for the fitted relations need not give us the reaction path of the economy at any time.

The present set of financial institutions and arrangements are quite different from those that ruled during the most recent financial panic-deep depression combination, that of 1929-33. Thus the question must be raised whether a financial panic of the sort that occurred in the past can now occur.

The hypothesis is that the financial panic which is present during deep depressions and absent during mild depressions is not a random exogenous affair; rather it is endogenous to the economy. The financial panic is made possible by the changes in the financial structure that take place during the long-swing expansion. As a result, the triggering event for a deep depression need not be specially severe, even though at an earlier date in the expansion a similar event could not trigger a deep depression.

Once a financial panic occurs, the decline in asset values, the forced changes in portfolios, and the revaluation of prospects combine to lower both consumption and investment demand, thus depressing income further than in a mild depression. We will not discuss the further ramifications of such a deflationary spiral; our attention will be upon the initiating event and the preparation of the environment for this event. However, it is worth noting that during the liquidation phases of a deep depression the financial "stage" is set for a long-wave expansion as debts are reduced, equity assets decline in value, and the stock of ultimate liquidity increases.

The view presented here is that the strictly defined money supply is a proxy for deeper, more significant financial variables; and that the observed relations between changes in the money supply and economic activity really reflect the impact of the evolving financial structure upon

<sup>3</sup> Ultimate liquidity can also be labeled outside money. It consists of government liabilities to the public (including the monetary system) and the gold stock.

economic activity. The central role of the money supply in past deep depressions is due to the institutional accident that commercial banks were the dominant financial intermediary. Under alternative institutional arrangements, where much greater weight is attached to non-bank financial institutions, a financial panic could occur even though the money supply does not fall. What changes do occur in the money supply may very well be induced by the behavior of the economy and the financial structure, witness the "excess reserves" of commercial banks in the 1930's.

Before we proceed, the financial structure of an economy should be defined. The financial structure of an economy consists of the set of financial assets and liabilities and the network of payment commitments among units. The financial structure includes the assets owned and the liabilities emitted by ultimate units, financial intermediaries, and governments. The "specialized" central bank guarantees, by organizations such as the FDC and the FHA, are a part of the financial structure. As a result of the existence of these organizations, it is now certain that in a crisis period certain losses will be absorbed by government and that an increase in the government debt and the money supply owned by the public will take place.

One view is that the institutional arrangements are now such that a financial panic cannot occur. Once again, it is maintained that we have entered a "new era." If the long swings are to persist and if the evolving financial structure is part of the generating mechanism for these swings, then the financial changes that accumulate over the expansion phase must be such as to constrain the sectors that would otherwise lead the continued expansion even in the absence of a dramatic financial panic.

The major task of this paper is to examine the implications of the cumulative changes in the financial structure that take place over a long-swing expansion for the stability of the financial system.

## II. *Financial Instability*

A. *The Determinants of Financial Instability.* Mild and deep depression cycles are, to a point, the result of a common cycle generating mechanism. However, in a deep depression cycle an initial downturn of income or a random decrease in particular asset values, triggers a general decline in asset values. In a mild depression cycle no such financial reaction takes place. Which type of cycle occurs at any time depends upon the stability properties of the existing financial system.

For the hypothesis central to this paper to be true, the changes in the structure of financial assets and liabilities that occur during a period in which only mild depression cycles take place must be such as to make the financial system less stable. Three dimensions of the financial systems will be emphasized here: (1) the rise of debts relative to income

for the income producing sectors, (2) the rise in price of stock market and real estate assets, and (3) the decrease in the relative size of ultimate liquidity.

The financial system is stable with respect to a change in income per period of a particular size or a given level of defaults per period on debts (or some combination of the two) if a change of this magnitude will not lead to a sharp fall in asset prices. Whether or not a financial system is stable with respect to changes of a given magnitude depends not only on the asset price, payments process, and ultimate liquidity factor mentioned above but also upon the extent to which government and central bank agencies absorb risk and maintain prices in various financial markets. Many of the government interventions in financial markets are of the form that will lead to an automatic improvement in the ultimate liquidity of the system and tend to stabilize the payments process once a financial panic is set off; there are automatic financial as well as income stabilizers.

Simple diagrams can be used to illustrate the stability properties of a financial structure. Let us assume that the initiating changes are declines in income and defaults in financial contracts, both as a rate per period. Given the existing financial structure there is a set of maximal shortfalls of income-financial defaults (both per period) combinations which cannot lead to a financial crisis; there also exists a set of minimal shortfalls of income-financial defaults per period which must lead to a financial crisis. Between these two border sets there is a set of income declines—financial defaults which may or may not lead to a financial panic. That is, for every financial structure the shortfalls of income-financial default plane is divided into regions in which the financial system is stable, unstable, and quasi-stable (illustrated in Chart I).

As the financial structure changes, the borders between the stable, quasi-stable, and unstable regions of the plane shift. The borders of the regions shift towards the origin as a long-swing expansion matures; that is, stable and quasi-stable regions become smaller parts of the plane.

Assume that the normal unamplified recession decline in income is in the range  $\Delta Y_1 - \Delta Y_2$  and the usual value of defaults per period, when no financial crisis is underway and income is falling by  $\Delta Y_1 - \Delta Y_2$ , lies between  $f_1$  and  $f_2$ . Two dimensions of the financial structure that change over the long swings will be used to illustrate the argument: the debt income ratios of households and business firms. For an initial change of the designated size the financial system will be stable, quasi-stable, or unstable depending upon the size of household and business debt income ratios; the higher the debt income ratio the less stable the financial structure (illustrated in Chart II).

In the light of the above, the hypothesis that the financial system evolves from being stable to being unstable over the long-swing expan-



The money on the wing perspective is not irrelevant for the current portfolio balance view of monetary theory. Money holdings yield services in kind in the form of protection against specific repercussions of various contingencies. The repercussions of not having enough money on hand are those which follow from not being able to meet payment commitments. Money is held at rest because of the conditions facing the owning unit which will lead to its taking wing. The size of the stock of money held relative to payment commitments depends upon the various arrangements available which generate assets that are "as good as money" for the satisfaction of precautionary motives, as well as the return from not holding money.

One aspect of the evolution of the financial system from being stable to being unstable is due to the growth of contractual payment commitments relative to both money holdings and specified money flows. In terms of standard symbols, the  $PT$  of Fisher's  $MV=PT$  should be broken down into components; that is

$$MV = \sum_{ij} P_{ij} T_{ij}$$

where  $i$  is the index for payment type and  $j$  is the index for classes of economic units. The argument is that over long-swing expansions the ratio's

$$\frac{P_{\alpha j} T_{\alpha j}}{P_{\beta j} T_{\beta j}}$$

change in a systematic fashion which tends to make the financial system less stable; in the language to be introduced  $\alpha$  will represent balance-sheet payments and  $\beta$  represents income payments.

Money is a means of payment. For each class of economic units, we can distinguish three classes of payment: income, balance-sheet, and portfolio payments. The classification of payment types and sectors is of course a construct imposed upon the world. As is true of any classification system there are observations which do not fall clearly into one or another of the cells and adjustments are necessary as the format is applied.

Of course each payment has two sides: a payment and a receipt; and a transaction need not be the same type to both sides.

The simplest way to begin examining the various types of payments is to look at the balance sheet of an economic unit and the commitments to make payments as stated in the liabilities. At any moment of time each economic unit has a balance sheet; the liabilities in the balance sheet finance a "position" as stated by the assets. Each liability carries with it a dated, demand, or contingent commitment to make payments.

Each asset, in the case of tangible assets, various collections of assets, generates receipts; in addition if necessary the economic unit is able to sell assets to acquire cash.

The labels dated, demand and contingent balance-sheet payments help explain the content of each class of balance-sheet payment. Dated payments act to constrain behavior; units will operate so as to have sufficient cash on hand at the designated dates to fulfill the commitments. Demand balance-sheet payments are of special importance for depository financial intermediaries, although various other units do emit demand or well-nigh demand liabilities. Units issuing demand liabilities must have a guaranteed refinancing source if they are to be viable under pressure. The government and central bank financial organizations have taken on contingent liabilities which do extend what amounts to guaranteed refinancing for many classes of financial intermediaries; although private endorsements and insurance contracts also are contingent liabilities. Some of the contingent liabilities are demographic in character, such as social security and pension fund liabilities; others are related to the behavior of the economy, such as acceptances, endorsements, and mortgage insurance.

The production of income involves both intermediate and final "income payments." Obviously, as interest payments due to debts are included among balance-sheet payments, the final income payments do not include all the items that enter into "income" as conventionally measured. Dividends on common stock are the result of current income production, and hence it can be argued that they should be included among income payments; however, as common stock is an asset in household and other balance sheets it may be best to include dividends as a special type of "contingent" balance-sheet payment in evaluating the changing structure of payments over time.

Portfolio payments occur when assets, real or financial, are traded. The "portfolio" aspects of trading in existing assets is obvious. However, in the creation and extinction of certain classes of assets the transactions can be a portfolio transaction for one and an income or balance-sheet transaction for the other party to the transaction.

For example, the withdrawal of a deposit from a savings and loan association is a balance-sheet payment by the savings and loan association and a portfolio receipt by the depositor. The payments made in the production of a capital asset are income receipts to the producing agents and portfolio payments by the firm or household that will own the real capital.

As a long-wave expansion progresses the relative weight of the different types of payments change: balance-sheet and portfolio payments rise relative to income, especially final income, payments. For the

income producing sectors, income receipts are the major source of funds for meeting payments due to liabilities. A rising balance-sheet payments-income receipts ratio for these sectors indicates that a given fall of income will force a larger proportion of units to try to obtain funds from other than their normal sources. These other sources are the sale of assets or alternatively "forced borrowing." That is, a "forced" portfolio transaction will need to take place as a result of a shortfall of income.

Portfolio sales will yield funds only if there are buyers (or, alternatively, lenders). If the order books are not thick relative to offers to sell, asset prices will fall sharply. Capital losses and a decrease in the protection offered debt owners by asset values will take place. Even units that were not initially affected will want more cash and risk free assets to protect themselves against becoming forced borrowers or sellers. Obviously a single unit's financial distress does not always lead to a financial panic; only if the initial financial situation is such that an income shortfall will affect many units and if asset prices are vulnerable to sharp decreases will this possibility exist.

Of particular importance for the development of a financial panic is the ability of financial institutions to meet these commitments and to continue acquiring assets. Financial distress for both bank or nonbank financial intermediaries affects both the ability of many units to make payments and the markets for assets. Hence "distress" for financial intermediaries seems to be necessary if a financial panic is to develop.

One element in the development of an "unstable" financial system is the rise in balance-sheet payments relative to final gross income for the income producing units. This will be measured by the ratio of debts to income for these units. A financial panic develops when a substantial number of units resort to portfolio sales to acquire cash, because the normal source of cash, income receipts, has not generated the expected amount. In particular, a financial panic can develop if financial intermediaries are forced to sell assets to acquire cash.

*C. Asset Prices.* During a period in which only short and shallow depressions occur, the economy will grow rapidly. As such a period increases in length, the market value of those assets that benefit from growth will rise to reflect not only the achieved growth but also expected growth. In addition, the value of protected private liabilities, such as bonds, mortgages, and deposits, will rise as the price of the underlying equity or real asset increases.

There is no "correct" price for a future income stream. The market price of an asset depends upon the time-path the income it yields is expected to follow as well as the certainty with which these expectations are held. As is well known, given that the capitalization rate for an

unchanging stream of earnings in a given risk class is

$$\frac{1}{r} \quad (r < 1),$$

the capitalization rate for a given stream of earnings in the same risk class that is expected to grow at  $g$  percent per year is

$$\frac{1}{r - g} \quad (g < r).$$

As growth expectations take over, during a long-swing expansion, the capitalization rate for assets that are expected to benefit from the growth in income will rise from

$$\frac{1}{r} \quad \text{to} \quad \frac{1}{r - g} \quad \text{times (current) income.}$$

The movement of asset values from

$$\frac{1}{r} \quad \text{to} \quad \frac{1}{r - g} \quad \text{times income}$$

will of course generate capital gains to holders of the asset, especially while the numerator is increasing at  $g$  percent per year. The "equilibrium" rate of growth of the price of these assets is  $g$  percent per year. However, while the expectations that growth will take place are being reflected in asset values, the price will grow at a rate faster than the equilibrium rate.

A purely speculative secondary run up of asset prices can occur as prices begin to reflect capital gains that occurred when the capitalization rate was increasing. This speculative secondary wave is inherently unstable, for once the rate of growth of asset prices slows down, the equilibrium capitalization rate, given growth expectations, of  $1/r - g$  becomes the determinant of asset values. The triggering event that breaks the secondary speculation wave can be a relatively slight shortfall of the rate of growth of income below the expected  $g$  percent per year.<sup>4</sup>

The implications of change in the capitalization rates and the secondary speculative rise in asset prices are of special importance for an

<sup>4</sup> It is worth noting that if the tax laws give a favored treatment to capital gains as against income, an annual income of  $\$1.00 + \frac{g}{r-g} (\$1.00 \text{ dividend}, \frac{g}{r-g} \text{ rise in market price})$  with a capitalization rate of  $\frac{1}{r-g}$  is worth more to the stockholder than an income of  $\$1.00$  with a capitalization rate of  $1/r$  even though both yield  $r$  percent per year on the asset price. That is, assets that generate a part of their returns in capital gains should have a smaller measured rate of return than assets which generate all of their returns in fully taxable income. This also argues for a rapid rate of increase in equity prices as growth expectations become generalized.

economy in which corporations are the dominant form of business organization and the ownership of corporate stock is widespread. The market price of real estate assets is also affected by growth expectations, and speculative waves in real estate prices occur.

The impact of speculative movements in real asset or common stock prices depends upon whether or not the inflated asset prices are fed back into the financial system as security for debts. The low margin stock market of the 1920's did feed assets whose prices reflected speculative factors into the balance sheets of financial institutions and non-financial corporations. The high margin stock market of the postwar period tends to minimize this feedback of common stock prices into the financial system, although the feedback from real estate prices and commodity transactions into financial institutions still takes place.<sup>5</sup>

D. *Ultimate Liquidity.* An economy's stock of ultimate liquidity consists of those assets whose nominal value is essentially fixed and which are not the liabilities of any private unit within the economy. The ultimate liquid assets carry no default risk and as they are essentially fixed in market value, they are always available to meet payment commitments. No private unit is constrained by payment commitments embodied in these assets.

In the United States the ultimate liquid assets consist of the gold stock, various types of treasury currency, and the public debt outside government trust funds. The inclusion of the government debt implies that if necessary support operations by the central bank will occur. A large part of the stock of ultimate liquidity is held by the monetary system. The extent to which the assets of the monetary system consist of ultimate liquid assets is a qualitative characteristic of the money supply. A growth of the money supply that reflects the growth of ultimately liquid assets owned by the monetary system has quite different implications for the behavior of the economy than a growth of the money supply that is due to the acquisition of private debts by the monetary system.

The ratio of the values of ultimately liquid assets and total financial assets in the economy is a measure of the extent to which financial positions are impervious to financial losses; the higher the ratio of ultimately liquid assets the more stable the financial system. In addition, the ratio of the value of ultimately liquid assets to income measures the extent to which a shortfall of income and its concomitant effect on asset prices will not result in a default of payment commitments.

Over a long-swing expansion the two major components of ultimate liquidity—the government debt and the gold supply—grow more

<sup>5</sup> Note that if the banking system grows less rapidly than income and the value of financial assets, and if nonfinancial business tends to decrease its dependence upon bank loans, a given proportion of stock purchases on margin can become a rising proportion of bank assets that reflect price sensitive assets.

slowly than income and other assets. As a result, the relative size of the stock of ultimate liquidity decreases, which tends to reduce the stability of the financial system. Of course, what happens to the size of the government debt is the result of policy decisions; therefore the decline in the relative size of the stock of ultimate liquidity need not take place over the long swing.

E. *Recapitulation.* A financial panic occurs when a not unusual decline in income or run of defaults on financial contracts occurs in a "favorable" environment. The favorable environment consists of (1) a high ratio of balance-sheet payments to income receipts for major classes of units, (2) equity and real estate asset values based upon growth expectations that reflect not alone the real growth potential of the economy but also the capital gains that occur as these assets are revalued to reflect growth expectations, and (3) a low ratio of ultimate liquidity to other financial assets and income. During a long-wave expansion each of these elements of the financial environment changes in such a manner as to increase the probability of a panic taking place; balance-sheet payments increase relative to income receipts, asset prices are bid up, and income and other financial assets grow faster than ultimate liquidity. Hence a financial panic is not something that just happens; it is an outcome of the very cyclical phase it brings to an end.

TABLE 1

RATES OF GROWTH OF FINANCIAL ASSETS AND LIABILITIES 1922-29, 1948-57 AND 1957-62

	GROWTH 29/22	RATES 57/48	%/Yr. 62/57
A. GNP current prices.....	5.2	6.1	4.5
B. Disposable personal income.....	5.0	5.6	4.4
C. Consumers and nonprofit sector: current receipts after deductions.....		5.5	4.5
D. Corporate nonfinancial business: gross profits after taxes.....	4.5	4.3	7.9
E. Market value of stocks listed in N.Y. Stock Ex- change.....	19.0	12.6	
F. Common stock: consumer and nonprofit organi- zation.....		11.7	10.2
G. Internally held federal debt.....	-4.3	0.22	1.85
H. Ultimate liquidity .....	-3.2	.15	1.50
I. Demand deposits net and currency.....	2.9	2.3	1.54
J. Time deposits at commercial banks.....	7.8	4.5	11.4
K. Total: demand+time deposits.....	4.8	3.5	3.7
L. Household and nonprofit sector ratio of total li- abilities to income.....	7.2	8.4	4.3
M. Corporate nonfinancial business ratio of total liabilities to income.....	0.3	3.8	-1.3

SOURCES: In general 1922-29 data are from R. W. Goldsmith, *A Study of Saving in the United States* (Princeton Univ. Press, 1955).

The 1948-62 data are from *Flow of Funds/Savings Accounts Supplement No. 5*, Board of Governors, Federal Reserve System, and various issues of the *Fed. Res. Bul.*

### III. *Conclusions*

The evidence presented in the attached table indicates that during 1922-29 and 1948-57 financial changes took place that were destabilizing. In particular, both the household liabilities-income ratio and common stock prices increased rapidly and ultimate liquidity either declined or grew very slowly.

The economy grew at a somewhat slower rate in 1957-62 than in 1922-29 or 1948-57. During this period of somewhat slower growth, on the whole the financial structure continued to evolve in such a way as to decrease its stability, albeit at a slower rate than in either 1922-29 or 1948-57. Household debts continued to grow relative to income, common stock prices continued to rise rapidly, and the stock of ultimate liquidity grew at a much lower rate than other financial instruments and income. The one really stabilizing change since 1957 has been the decline in the ratio of corporate nonfinancial business liabilities to income.

This paper contains no answer to the questions as to whether or not a financial panic followed by a deep depression can now occur or whether a long-wave contraction can take place in the absence of a financial panic. The barriers to a financial panic erected in the aftermath of the great crash have not been tested; all we can assert is that during the period of slower growth since 1957 the destabilizing financial changes have continued at a slower rate. Neither an unwinding of the financial changes of 1948-57 nor a convergence to a steady, equal rate of growth for the various dimensions of the financial system has taken place.

## DISCUSSION

JAMES DUESENBERY: For once I find myself discussing papers which I can wholeheartedly approve. It is easier to criticise than to express admiration. Let me simply say that I wish I had written both of these papers.

Mr. de Leeuw's paper is a very impressive one. He not only displays a mastery of monetary theory and statistical technique but he shows considerable artistry in the construction of his model. He has managed to include in his model the main institutional features of our complex financial system, while keeping the scale of his model down to manageable proportions. Anyone who has tried his hand at realistic financial analysis will realize that that is in itself a considerable achievement.

At the same time, his model reflects the empirical content of a number of the main theoretical developments in monetary analysis. His equations reflect the recent emphasis on the "portfolio balance" approach to monetary theory, though they also take account of lags in the stock adjustment process and the importance of current flows as factors which influence the speed of adjustment. De Leeuw's demand for money equations reflect the lagged adaptation or permanent income approach suggested by Friedman. When combined with the term structure equation, they also include the "speculative" demand for money so much emphasized by Keynes. In addition, the demand for money equations show the substitutions between savings deposits and demand deposits and currency emphasized by Shaw and Gurley and others.

In that connection, though, the substitutability of liquid assets for one another would appear more clearly if the demand for nonbank thrift deposits—mutual savings bank and savings and loan accounts—were shown separately instead of being combined with insurance and pensions.

Mr. de Leeuw's paper gives a good demonstration of the power of the technique of simulation. Simulation is a substitute for the mathematical analysis of the implications of a set of hypotheses about the response of one economic variable to the movements of others. When we use a mathematical analysis we cannot get concrete results about the nature of dynamic processes except for relatively simple models. A model which can be analyzed mathematically must simplify reality to such an extent that it must distort the facts or discard consideration of important aspects of the system. We can work out the concrete implications of much more complex systems by means of simulation.

That is particularly important in an area containing a number of policy parameters which separately influence subsectors of the system. Changes in the maximum interest rate paid on commercial bank time deposits appear to have had significant effects on the demand for money. But developments of that sort can only be taken into account by a rather detailed model which is too complex for mathematical analysis but which can be analyzed by simulation.

Mr. de Leeuw does not pretend that his model represents final truths in



the financial field. Some further disaggregation will undoubtedly be necessary to separate out the thrift institutions as suggested above, to separate business from household demand for financial assets, and to separate the demand for bank loans by certain types of customers from the rest. It may also be necessary to develop equations for a separate mortgage market. At the same time, there are alternative hypotheses about some of the variables. Broad time series data do not give a sufficiently sharp test of alternative hypotheses. A long testing program using more micro data will be necessary before we are completely satisfied with the model.

Nonetheless, I feel that Mr. de Leeuw's work has brought us a long step forward in the empirical analysis of the financial system.

Professor Minsky's paper is an equally interesting one, though entirely different both in style and content. He argues that the ratio of debt to income tends to rise progressively during a long period of prosperity interrupted only by minor recessions. The increase in debt ratios may operate as a drag on the rate of investment and slow down or bring to a halt the growth of demand. Of course, the system may simply move toward an equilibrium debt income ratio and continue to grow to scale. The effects of debt on the system are analagous to those of accumulating capital in simple capital-stock-adjustment models.

But even if the system is capable of growing at some equilibrium rate with corresponding equilibrium debt-income and capital-stock-income ratios, it is more vulnerable to shock at high debt-income ratios than at low ones. A decline in income from some minor cause might have only minor secondary effects in the early stages of a long upswing. At a later stage, when debt income ratios are higher, a similar event might set off a downward spiral of decreasing investment and income, with bankruptcy and fear of bankruptcy spreading throughout the system. The possibility of that happening is increased when debtors and creditors are willing to accept high debt-income ratios only because of the optimism engendered by a long period of prosperity.

Professor Minsky mentions the long-cycle hypothesis but it is not necessary to believe in any regular long cycles to believe that debt income ratios do tend to increase as the length of the boom increases and that such increases render the system increasingly vulnerable to shock.

In his paper Minsky emphasizes the overall increase in debt ratios. But it is important to remember that there is always wide dispersion in the growth rates of particular sectors. Historically, some particular rapid growth sector has often become particularly vulnerable long before any dangerous situation developed in the aggregate debt position.

Finally, I should like to note my agreement with Professor Minsky's comments on stock market vulnerability. I am in general agreement with the spirit of Professor Minsky's paper and I am glad that he has again drawn our attention to a class of problems which has received too little attention recently.

Before concluding, I would like to draw your attention to an important methodological problem posed by these papers. Mr. de Leeuw's model is

based on the empirical regularities connecting the short-run movements of financial variables in a given institutional setting. Professor Minsky emphasizes the importance of the evolution of financial institutions and markets. He also emphasizes the importance of slowly moving processes—like debt accumulation, which reveal their effects only in response to a shock whose timing may be accidental. Professor Minsky's fears cannot be readily reflected in a model like Mr. de Leeuw's. One of the most important and vexing methodological problems is that of finding ways to construct empirical models which do full justice to both kinds of consideration.

We can feel gratified by the progress in financial analysis which these papers reflect but it is also clear that we have a long way to go.

JOHN G. GURLEY: Any economic system, or almost any, has in it real things and financial things: structures, equipment, land on one side and bonds, deposits, and currency on the other—real wealth at the base and a financial superstructure over it.

These two papers approach this economic edifice in remarkably different ways. Mr. de Leeuw's paper is primarily concerned with how real things affect financial things (and how some financial things affect other financial things); Professor Minsky seems to care more about how financial things influence real things. Moreover, De Leeuw is concerned with a short period of time; Minsky's eye sweeps over decades and even centuries. Again, De Leeuw's approach is elegant, sparkling with two-stage least squares,  $R$  squares, homogeneity postulates, and plenty of  $t$  values. Minsky's approach is by comparison vulgar: he speculates, tosses out wild suggestions, and tests practically nothing. Worse yet, Minsky's paper is not even pure vulgarity, for at the very end, perhaps for the sake of convention, he puts together a few figures, some relevant, some great surprises. Finally, De Leeuw's paper is not only sophisticated but clearly written; in contrast, Minsky failed to enunciate clearly into the dictaphone, forcing his typist to make up several dozen words, which she apparently placed at random in the blank spaces. Many hilarious sentences were born of this procedure, and this makes the reading of Minsky's paper—the copy I have, anyway—one of the joys of the entire meeting.

I hope that I have not disguised how much I really like Professor Minsky's paper—not just for its funny, garbled sentences, but also for its ideas. Most of us, I suppose, have long ago put financial panics out of our minds, so now it does seem a little quaint to hear someone seriously discussing them. But Professor Minsky argues pretty persuasively that financial panics are not necessarily things of the past. In fact, his figures indicate that during recent years the financial superstructure has grown in much the same threatening way it did during the 1920's. First, private debts have grown very fast relative to private incomes. Second, stock prices have soared. Third, ultimate liquidity has become a smaller and smaller part of the financial superstructure.

If I may use some of my own figures, the private debt-income ratio has risen from around 0.9 in the early postwar years to 1.5 today. Stock prices

have about quadrupled. And ultimate liquidity has fallen from 30 percent of total financial assets to around 10 percent.

Minsky is correct in stating that this is not unlike the 1920's. And I share his apparent feeling that another financial panic is quite possible. Still, it seems to me that he fails to stress adequately the difference between 1929 and the present. At that time, the federal government was backing, in one way or another, only 4 percent of total financial assets. Today, federal government securities, issues of federal financial intermediaries, insured and guaranteed primary securities, such as mortgages, insured and guaranteed indirect securities, such as bank deposits—all of these add up to somewhat more than 25 percent of total financial assets. This is quite a difference, but it should be noted that this ratio has been falling steadily since 1945, when it was 42 percent; 25 percent is a lot of federal backing, but it is a far cry from the backing we had fifteen years ago. Meanwhile, the trends, which would appear to weaken the superstructure, go on; and it is a very pertinent question to ask, as Professor Minsky does, whether a future financial crisis is now in the making.

Mr. de Leeuw's financial model is an ingenious one; it shows a high level of technical competence in addition to a thorough knowledge of financial markets.

His model is one part of the complete S.S.R.C. model of the economy. Consequently, he takes many real things as given in his attempt to explain some key financial variables. But this does not make his task an easy one, for he has chosen to consider, as most investigators have not, a wide array of financial variables and interest rates. His model is based on demand-supply conditions in four markets for indirect debt (currency, demand deposits, time deposits, and savings and insurance claims), two markets for primary securities (U.S. government securities and private loans and securities), and the market for bank reserves. These markets comprise a large part of our financial structure, and Mr. de Leeuw has done very well to bring them together so neatly in one package.

De Leeuw's behavioral equations have an interesting form, which I can illustrate best by comparing his demand-for-money function with a standard one. The latter shows the real demand for money (nominal demand divided by the price level) rising with real income, moving opposite to the rate of interest, and rising with a real wealth variable. (In Patinkin's formulation, real wealth is simply the real amount of outside money—all other claims cancelling out and real assets neglected.)

De Leeuw departs from this formulation in several ways. First, he chooses to focus on the change in demand for money rather than its absolute amount. Second, he divides all dollar variables by a wealth variable, so that his dependent variable becomes the ratio of the change in money demanded to wealth. Third, he deals in nominal, not real, variables: the absolute price level is not separately considered; in this model, financial markets determine relative prices (i.e., interest rates) and not the absolute price level. Fourth, De Leeuw explains the change in demand for money with several interest rates, not just one; and with several versions of income (and spending).

Fifth, a key explanation of the change in demand for money is its absolute amount in the previous quarter. And, sixth, money is divided into its two components of currency and demand deposits, and De Leeuw sets out to explain changes in each.

The other behavioral equations have very much the same characteristics, and there is a long string of such equations. All in all, the model is so very complicated that one can see some of its properties only by plugging it into a computer, or perhaps by putting it under one's pillow for several nights.

One writer, in his recent review of an encyclopedia, said that "evaluating a new encyclopedia is like reporting on the success of a marriage; one cannot tell much after living with the work for only a month." I feel much the same way about De Leeuw's model: I would like to live with her for awhile before definitely saying "yes."

But, returning to the real model, my initial impression is that the machinery is terribly elaborate and complex for the stuff it grinds out. That is, for the most part, the results of the simulations appear to me to be rather uninteresting. Interest rates fall when the money supply is increased; banks buy more earning assets when reserve requirements are lowered; time deposits increase during a recession; and so on. If this complicated mechanism can turn out answers like that, it simply must have some attractive features.

I am fairly sure that it has. The model represents a most interesting attempt to bring together a host of financial phenomena, and De Leeuw should be congratulated on the outcome.

ALLAN H. MELTZER: The papers at this session provide an opportunity to contrast the procedures that economists formerly employed with those that are currently in vogue. De Leeuw's paper is an outstanding counter example to the notion that substantive empirical work is incompatible with a paycheck from the Board of Governors. He uses many of the new "sophisticated" techniques that are currently the vogue among econometricians. There are nineteen formally stated hypotheses, simultaneous equation estimates, single equation estimates, and, as a capstone, a simulation of the whole system. In contrast, Minsky's paper provides a general discussion of the role of financial institutions in business cycles, revives the long wave—that I thought had been buried—but provides no clearly stated, empirically supported hypotheses and nothing that can be seriously regarded as evidence.

There is no question about which procedure adds most to our knowledge. The De Leeuw paper tells us much about the correspondence of a particular set of hypotheses to a particular set of data. As such, it provides much evidence to support the general framework of a process transmitting the impulse of monetary policy action, or the behavior of the public, including financial institutions, through the credit markets. But we learn only that, in general, De Leeuw's hypotheses perform better than chance—the only alternative considered.

I suggest that it is time that economists stop regarding chance as the only

alternative. Sufficient empirical work has been done in the past ten years in the areas of supply and demand theory for money or in the behavior of financial markets to indicate that there are better alternatives than chance. I do not see how we can make progress if each of us proceeds by constructing his own set of hypotheses and largely ignoring the work of others.

Much more than the question of footnotes and bibliography is involved. It is the comparison of alternative hypotheses that is at issue. We cannot hope to reject particular hypotheses and concentrate on improving others if we do not test hypotheses against alternatives rather than against a null hypothesis; i.e., I do not see how we can improve our knowledge about monetary or other economic processes unless we systematically compare alternative hypotheses using the same time periods and the same data where applicable.

Our growing knowledge of single equation and simultaneous equation estimates bears on this point. From the studies that I have seen, and from the work that I have done with Karl Brunner, I have become convinced that single equation bias is generally an insignificant problem relative to single model bias. De Leeuw's paper adds further evidence on this point. His single equation and simultaneous equation estimates do not differ markedly. What we do not know, because we rarely ask, is the answer to the question: How much better is the proposed model than other verifiable models purporting to predict and describe the same events? Only when we start to answer that question can we begin to develop systematic knowledge that can form the basis of rational policy judgments.

The single model bias to which I refer is not a technical statistical problem; it is a problem in economics. The point, though often neglected, is quite simple. I am arguing, first, that we should choose among alternative hypotheses by comparing them systematically and, second, that we begin to derive our models from and relate our findings to the corpus of economic theory so that the validated stock of knowledge may grow.

Some specific examples from De Leeuw's work may help to clarify the point. My work with Karl Brunner,<sup>1</sup> to which De Leeuw refers, suggests thus far, that if the demand function for money is not specified as a part of the general wealth adjustment process, the estimated interest elasticity of the demand for money is too low to explain both the observations for the postwar period and those for time periods going back to the late nineteenth century. De Leeuw's estimated interest elasticity is very similar to those obtained from the Latané form of the general Keynesian model. I suggest that his model is misspecified. But I cannot assert with complete confidence that Brunner-Meltzer are right and that De Leeuw is wrong. However, I do assert that comparison of alternative demand equations will have an appreciably greater effect on the resolution of this question than concern with so-called "more sophisticated" statistical procedures.

I will draw a second example from the many available. Meigs's study of

<sup>1</sup> "Predicting Velocity: . . .," *J. of Fin.*, May, 1963. I hasten to note that other demand functions may produce this result in the future. None have to date.

the demand by banks for free reserves<sup>2</sup> appears to be at least as powerful in explaining monthly variations in the demand for free reserves as De Leeuw's quarterly demand equation for excess reserves and borrowing. De Leeuw acknowledges the contribution by Meigs's but he does not explain why he modified the Meigs's hypothesis. How can we expect to develop validated economic theory if we discard useful, partially validated results without showing that we have better hypotheses to take their place?

Before leaving the new look and turning to the old, let me add that the De Leeuw model has numerous interesting features. I would be happy to learn more about the reasoning underlying the particular hypotheses. But I would suggest to the author that before we attempt appraisals of policy, as the author does in his simulation study, we ought to be sure that we are basing conclusions on the best available model, not on the one that uses the newest techniques.<sup>3</sup> That the author agrees in his conclusion with the need for detailed comparisons of existing alternative hypotheses only reinforces my view that the single model bias ought to receive at least as much attention as we currently lavish on the single equation bias.

I cannot understand or accept the position from which Hy Minsky starts, viz., that we take the long wave as given and explain the development of financial institutions. I should think it more fruitful to explain why prices or income fall or rise. And, unless I am badly informed, there is general agreement that monetary factors play a not insignificant role in the explanation, at least for long-run price movements.

But my principal comment on Minsky's paper is to question his explanation of the role of financial factors in deep depression and the suggestion that these are summarized in some vague way by the ratio of liabilities to income. It is easy to show that his ratio is nothing more than a reflection of asset yields, interest rates, and relative prices, if we write the equations for the balance-sheet and income statement identities.<sup>4</sup>

$$(1) \quad W_h + A - L = W_n + W_h = W$$

$$(2) \quad Y_h + RA - rL = Y_n + Y_h = Y$$

from which it follows that the reciprocal of the Minsky ratio

$$\frac{Y}{L} = \frac{Y_h + RA}{A - W_n} - r.$$

The equations make clear what Minsky's paper does not: Whether the

<sup>2</sup> A. J. Meigs, *Free Reserves and the Money Supply* (Univ. of Chicago Press, 1962).

<sup>3</sup> I might add that de Leeuw's conclusion that the money multiplier is much lower than the reciprocal of the weighted average reserve requirement was anticipated by Karl Brunner for the U.S. in "A Schema for the Supply Theory of Money," *Int. Econ. Rev.*, 1961. The same paper explicitly notes that open market operations and changes in reserve requirements have equal effects on the supply of money. My own work for France (*J.P.E.*, 1959) finds that open market operations in the French system have a multiplier effect of approximately 2 on the money supply despite the absence of formal reserve requirements.

<sup>4</sup> Symbols are defined as follows:  $W_h$ =human wealth;  $W_n$ =nonhuman wealth;  $W$ =total wealth;  $A$ =total assets;  $L$ =total liabilities;  $Y_h$ =income from human wealth;  $R$ =the gross yield on asset portfolios;  $r$ =the cost per dollar per time associated with the consolidated debts of the private sector;  $Y_n$ =net income from nonhuman wealth;  $Y$ =total income.

ratio of income to liabilities is high or low is a matter of relative prices and interest rates that are determined by the interaction of markets for real and financial assets and the services of human and nonhuman wealth. It is most unlikely that we can short-cut the process of investigating the determinants of these yields and prices in the context of a verifiable model connecting stocks and flows. Indeed, it was because of the failure to do more than learnedly discuss such problems that the old economics gave way to the new.

Whether the money supply is a proxy for deeper, more significant financial variables, as Minsky states, we will learn only if we attempt to test—i.e., compare—explicit alternative hypotheses. But I think that we can say with confidence that the position of the banking system is not an institutional accident. It is a substantive issue that can be explained in terms of hypotheses on which Minsky commented at a similar meeting last year.<sup>5</sup>

<sup>5</sup> Brunner and Meltzer, "The Place of Financial Intermediaries in the Transmission of Monetary Policy," *A.E.R.*, May, 1963.

## REAPPRAISALS IN AMERICAN ECONOMIC HISTORY

### A NEW LOOK AT HUNTER'S HYPOTHESIS ABOUT THE ANTEBELLUM IRON INDUSTRY

By PETER TEMIN  
*Harvard University*

The process whereby innovations diffuse throughout an economy or throughout the world has interested many people at many times. One of the more intriguing examples of this process is the spread of the use of coke to make pig iron, which has interested investigators because of two seemingly incompatible properties. The use of coke in the blast furnace was one of the enabling innovations of the industrial revolution in Britain, and it played a similarly important role in the industrialization of many, if not all, other industrializing countries in the nineteenth century. Despite this importance, however, the spread of coke was not rapid; there was often a delay of as much as a half a century in the spread of this innovation from Britain to other countries. Coke was the fuel for well over 90 percent of the blast furnaces in Great Britain before 1810.<sup>1</sup> But half a century later, at the start of the Civil War, the proportion of pig iron smelted with this fuel in the United States had barely attained the level of 10 percent of the total.<sup>2</sup>

The example of America is particularly intriguing. The contrast between the long delay in adopting the new process and the subsequent rapid expansion of its use is striking.<sup>3</sup> In addition, the delay has been made to appear even more problematical by the recent book of Habakkuk, which has emphasized the extent to which the United States led Britain technologically in certain areas before the Civil War.<sup>4</sup> The problem before us is the reconciliation of the evidences of continuous progress in the United States with the neglect of this important innovation.

Many reasons have been given to explain the half-century delay in the adoption of coke in America. The strongest reasons given in the nineteenth century were that charcoal was much more plentiful in

<sup>1</sup> Thomas S. Ashton, *Iron and Steel in the Industrial Revolution* (Manchester: Univ. Press, 1924), p. 99.

<sup>2</sup> James M. Swank, *History of the Manufacture of Iron in All Ages* (2d ed.; Philadelphia, 1892), p. 376.

<sup>3</sup> *Ibid.*

<sup>4</sup> H. J. Habakkuk, *American and British Technology in the Nineteenth Century* (Cambridge: Univ. Press, 1962).



this country than in England and that American ironmasters were opposed to the use of coke due to ignorance, irrationality, or both.<sup>5</sup> Louis C. Hunter, in a classic article published in 1929, showed conclusively that these reasons were false and propounded a new hypothesis to explain the delay.<sup>6</sup> This paper will build upon Hunter's masterful analysis and suggest some alterations in his hypothesis about the antebellum iron industry.

Hunter pointed out that the cost of charcoal was almost entirely the cost of labor incurred in preparing it, the cost of the original wood accounting for little or nothing in the price. The availability of wood in the United States consequently had little influence on this price, and charcoal—here as in England—was a more expensive fuel than coke. As for ignorance and irrationality, Hunter showed that the pig iron made with coke was inferior to the iron made with charcoal, which implies that the reluctance of the ironmasters to accept a new fuel in place of the old was an accurate, rational reaction to quality differences in the iron produced. These two observations by Hunter form the cornerstone of any modern discussion of the diffusion of coke smelting in America.

Hunter did not spell out the mechanism by which the low quality of pig iron made with coke restricted its production, but the obvious inference is that the low quality implied a low price which destroyed the profitability of the new process. Pig iron made with coke sold for a 20 percent discount from the price of charcoal pig iron during the 1850's, a difference of six dollars. The cost of charcoal to make a ton of pig iron was approximately nine dollars, and I have endeavored to show elsewhere that the reduction in the price of fuel attendant upon the use of coke was not sufficient to compensate for the reduced price of the product.<sup>7</sup> The antebellum ironmaster was offered no inducement in the form of higher profits to use coke—a fact which is borne out in the not completely pleasant experiences of those ironmasters who tried. There were good reasons, in other words, why the technical progress in the antebellum era noted by Habakkuk for light manufacturing did not extend to all areas of the economy.

By the advent of the Civil War, however, a change was beginning to be evident, and the amount of pig iron made with coke began to rise faster than the total production of pig iron in America. The second part of Hunter's hypothesis endeavors to explain this development. It states that the character of the demand for iron was changing at

<sup>5</sup> Swank, p. 366.

<sup>6</sup> Louis C. Hunter, "The Influence of the Market upon Technique in the Iron Industry in Western Pennsylvania Up to 1860," *J. of Econ. and Bus. Hist.*, Feb., 1929, pp. 241-81.

<sup>7</sup> Peter Temin, "A History of the American Iron and Steel Industry from 1830 to 1900" (unpublished doctoral dissertation, M.I.T., 1964), Chap. III.

the time of the Civil War, from a price-inelastic demand that was sensitive to changes in quality to one that was more price-elastic and less sensitive to changes in quality. The former pattern was characteristic of an agricultural demand in which small lots of iron were bought for use under a wide variety of conditions; the latter was the result of an industrial demand in which larger amounts of iron—relative to the total purchases of a firm or farm—were purchased for a specific purpose. The most important specific purpose was for the production of rails—a product that did not benefit from the use of high quality iron.

Two factors urge the reconsideration of this part of Hunter's hypothesis. As with the earlier part of the argument, Hunter did not outline a mechanism by which the change in demand caused a change in the method of production. Such a mechanism is considerably harder to specify than the one which translated a low quality of product into low profits because it involves a shift in demand curves, and even a shift in the elasticity of demand curves—always a difficult event to identify. In addition, recent work on the proportion of pig iron used for the production of rails has raised the possibility that the shift in demand noticed by Hunter was not large enough to effect the change in productive techniques under discussion.<sup>8</sup>

I would like to suggest an alternative to this part of Hunter's hypothesis. In the course of discriminating between my suggestion and the theory proposed by Hunter, both the mechanism involved and the relation of this mechanism to other work will be shown.

The increasing use of coke for a blast furnace fuel at the time of the Civil War may be seen as a reaction to changes in the cost and price structure of the American iron industry. Hunter's hypothesis is that the changes were in the demand for iron, that the demand for low-quality, low-cost iron rose relative to the demand for high-cost and quality iron. An alternative explanation is that the cost of making coke pig iron fell or that its quality improved; i.e., that the changes were on the supply side. The cost of making pig iron with coke does not seem to have changed with respect to the cost of using charcoal before the 1870's and the introduction of "hard driving," and this explanation may be dismissed. Discrimination between the two remaining alternatives—Hunter's hypothesis of changing demand and my suggestion of quality improvement in coke pig iron—is a harder task.

This task is facilitated by the existence of a third fuel—different from both charcoal and coke—suitable for use in the blast furnace and a third type of iron. The fuel was anthracite, found in eastern Pennsylvania and parts of Wales. Coke was used before anthracite in Britain, as the technical knowledge necessary for the ignition of the somewhat

<sup>8</sup> See the work of Fogel cited below.

recalcitrant anthracite was not discovered until the late 1830's, but anthracite became the variety of mineral fuel used initially in America. Its production rose rapidly in the 1840's, and it accounted for over half of American pig iron production in 1860.<sup>9</sup>

There are two possible reasons why anthracite was adopted before the Civil War and coke was not, corresponding to the two reasons proposed to explain why coke was adopted later. Anthracite is found east of the Alleghenies, while bituminous coal lies to the west. Demand may have differed in the two regions in such a fashion as to induce the production of a type of pig iron in the East that could not have been profitably produced in the West. Alternatively, the quality of the pig iron made with anthracite may have differed from that of coke pig iron, while the character of demand was unimportant. (The costs of making pig iron with the two different mineral fuels were approximately equal at this time. If there was a difference between them, it was that it was cheaper to produce pig iron with coke than with anthracite.)

As long as the commerce in heavy products such as pig iron was separated by the Allegheny Mountains into two markets, there was little evidence generated that would enable the modern observer to discriminate between these hypotheses. In 1852, the Pennsylvania Railroad completed its through line from Philadelphia to Pittsburgh, and the cost of transportation across the mountains was reduced. Pig iron made in eastern Pennsylvania with anthracite began to be sold in Pittsburgh in competition with the locally produced pig iron made with coke. If the character of demand was different in the two regions, we would expect pig iron made with anthracite to sell in Pittsburgh at approximately the price of pig iron made with coke; i.e., at a substantial discount from the price of pig iron made with charcoal. If, on the other hand, it was the nature of the varieties of pig iron that made a difference, we would expect to find an equally clear price differential existing between pig iron made with anthracite and with coke.

As it turns out, pig iron made with anthracite sold in Pittsburgh at the same price as charcoal pig iron.<sup>10</sup> Consumers in the West were willing to pay substantially more for pig iron made with anthracite than for pig iron made with coke in the antebellum era, and we conclude that pig iron made with anthracite differed significantly in its quality from pig iron made with coke. The reasons for this are clear. The main debilitating element in the coking coal available in the 1840's and 1850's was sulphur, and the anthracite deposits of eastern Pennsylvania were relatively free of this impurity.

The delay in the use of coke in the United States may now be placed

<sup>9</sup> Temin, Appendix A; Swank, p. 376.

<sup>10</sup> Hunter, "A Study of the Iron Industry of Pittsburgh Before 1860" (unpublished doctoral dissertation, Harvard, 1928), pp. 393-433.

in perspective. Ironmasters in the United States were adopting the new technology based on mineral fuel. Because of a difference in resources, however, they adopted it in a different form than that used originally in Britain. And because the innovations permitting use of the American resources, i.e., anthracite coal, were not introduced until the 1830's, they adopted it after a delay. The deposits of anthracite in the United States are very highly concentrated in a few counties of northeastern Pennsylvania, and the high cost of transport to the far reaches of the large American economy prevented the use of this fuel from completely eclipsing pig iron production with charcoal. Consequently, pig iron made with charcoal still accounted for about a third of the total at the start of the Civil War.

By this time, the cost of transportation had declined enough to permit pig iron made with anthracite to compete with pig iron made with charcoal in places other than the eastern seaboard. But as Hunter noted, more changes were in progress at the close of the antebellum years than those in the cost of transport. In a closely related development, the demand for iron was beginning to shift. The plentiful use of castings typical of the 1840's had begun to wane; the demand for rails had begun to rise.<sup>11</sup> The economy was also expanding geographically, and this led in turn to the discovery of new resources. We seek to know which of these developments made the use of coke in western blast furnaces more attractive than the importation of pig iron made with anthracite from eastern Pennsylvania.

Price comparisons again carry the key. The price of pig iron made with coke had to rise relative to the price of other types of pig iron if coke rather than one of the other fuels was to be used. If this price rise occurred because of a change in demand, as Hunter asserted, a price differential should have been maintained between pig iron made with anthracite and pig iron made with coke. The former class of iron was of higher quality than the latter before the Civil War, and there is no evidence that its quality deteriorated after that time. Only if the quality of pig iron made with coke improved until it was equal to that of pig iron made with anthracite could their prices become equal.

Yet this is precisely what happened. The extant price data for the Civil War years are obscure due to the paucity of comparable sources and the rapid changes in prices in these years. When prices do become available on a comparable and sustained basis, in the early 1870's, they show that pig iron made with coke and pig iron made with anthracite sold for the same price. In fact, the popular method of quoting prices was to refer to pig iron by grade, including iron made with anthracite and with coke in a single classification.<sup>12</sup> The quality of

<sup>11</sup> Temin, Chaps. I and II.

<sup>12</sup> The first regular market report appeared in the *Engineering and Mining J.*, 1874, p. 55.

pig iron made with coke had improved to the point where this type of iron was interchangeable with pig iron made with anthracite. There had been a change in the supply curve of pig iron made with coke, in other words—a change that enabled producers of pig iron with this fuel to supply more value to consumers with the same expenditure.

To what can we attribute this change in supply? The exploitation of high-quality coking coal is the obvious candidate. The modern observer of the antebellum iron industry cannot help but be struck by the absence of one of the most famous names of the American iron industry. I refer to Connellsville, the home of the metallurgical coke known by its name and famous for both its hardy physical structure and its relative chemical purity. This variety of coking coal was used in the expansion of coke pig iron production after the Civil War; its freedom from sulphur may be taken as the cause of the improvement in the quality of pig iron made with coke. The Connellsville coal region was initially discovered in the 1840's, but its exploitation is usually dated from 1859—the date of the construction of the first blast furnace in Pittsburgh designed specifically to use this coal.<sup>23</sup> This deposit was newly discovered in the boom of the 1840's and did not become known in the few years of the boom. The following decade did not witness a rise of pig iron production above its previous peak, and the inducement to use new resources was not strong. Only when the production of pig iron began to rise again at the end of the antebellum era was the new source of coke exploited.

While this provides an explanation for the improvement in the quality of pig iron made with coke, it is not clear that the discovery of new coal fields can be made a simple function of the expansion of the economy. This was certainly a potent factor, but it is also possible that ironmasters increased their search for new coal on the basis of increasing technical knowledge. Quite possibly, western ironmasters only came to realize that poor coal was the source of their difficulties in the course of the 1850's, their lack of exploration previously being determined as much by ignorance of what to look for as ignorance of where to look. We find an awareness of the difficulties of using sulphurous coal in these years, but we also find early writers criticized by later ones for their overoptimism on the usefulness of many coal deposits.<sup>24</sup> The process by which Connellsville coke was brought into use, therefore, retains some of its mystery even now.

---

It assumed its complete form a month later, in *ibid.*, 1874, pp. 134-35, when the two types of mineral fuel pig iron were classified together in the Pittsburgh report. An earlier report of sales appeared in *Iron Age*, 1873, p. 19. It showed the prices of the two types of mineral fuel pig iron at about the same level.

<sup>23</sup> Swank, pp. 476-77; Howard N. Eavenson, "The Early History of the Pittsburgh Coal Bed," *Western Pennsylvania Hist. Mag.*, Sept., 1939, pp. 165-76.

<sup>24</sup> Frederick Overman, *The Manufacture of Iron* (3d ed.; Philadelphia, 1854), p. 130; J. P. Lesley, *A Manual of Coal and Its Topography* (Philadelphia, 1856), p. 29.

The changing supply curve for pig iron made with coke thus explains the increased use of coke in this country. When a supply of bituminous coal became available that would produce an iron equal in quality to that made with anthracite, the relative price of coke pig iron rose and its production was encouraged. The growth of the transportation network, therefore, encouraged the universal use of mineral fuel in the United States iron industry through promoting geographical exploration and the exploitation of new resources, rather than by improving the competitive position of anthracite pig iron, as might have been anticipated in 1850. In addition, technical developments after the Civil War reduced the cost of using coke relative to the cost of using anthracite, and coke came to be the universal fuel for American blast furnaces.<sup>15</sup>

But what about the change in demand noticed by Hunter? If the preceding argument is correct, the switch to coke would have taken place whether or not there was a change in the character of the demand for pig iron around 1860. Such a change, however, could have helped the transition, and we may ask whether or not it did.

The most important component of demand from this point of view is the demand for rails, as the quality of iron used was not a major concern in this area. Robert Fogel has estimated that the proportion of pig iron used for the production of rails in the antebellum era was less than 10 percent.<sup>16</sup> This low figure, coupled with the absence of a rising trend near 1860, suggests that the effects of changing demand may be neglected. Albert Fishlow has revised Fogel's estimates, and his data show that the proportion of pig iron used for rails was rising in the 1850's, reaching a level of about 20 percent by the end of the decade.<sup>17</sup> This estimate gives greater scope to the influence of demand, but even it does not say anything about the actual effect.

The problem is that even if the production of rails did use an increasing proportion of the pig iron produced in the late 1850's, this could have been the effect of an increased supply of low-cost pig iron—made with coke—as easily as it could have been the cause of the increased production of pig iron with coke. The resolution of this problem requires the specification of the relevant demand and supply functions. Among the demand functions that need to be specified is

<sup>15</sup> Temin, Chaps. VII and IX and Appendix B.

<sup>16</sup> Robert W. Fogel, "Railroads and American Economic Growth: Essays in Econometric History" (unpublished doctoral dissertation, Johns Hopkins, 1963), Chap. V.

<sup>17</sup> Albert Fishlow, "The Economic Contribution of American Railroads Before the Civil War" (unpublished doctoral dissertation, Harvard, 1963), Chap. III. Fishlow also estimates the nonrail demand for iron coming from the railroads. While the magnitude of the total railroad demand for iron is important for other purposes, it is not relevant to a discussion of technological change in the iron industry. Many of the products used by the railroads did not differ from products used elsewhere in the nature of their demand for iron, and the volume of their production may be neglected.

the demand for rails, including the demand for rails to be used to replace worn-out rails, and the nature of this demand (together with the reciprocal supply of scrap coming from the worn-out rails themselves) is precisely the point on which Fogel and Fishlow differ. The problem does not appear to be easily soluble, and at this point we cannot assess the extent to which changes in the composition of demand helped the transition from charcoal to coke in the blast furnaces of the American iron industry. What the preceding argument has shown is that this help was not needed to effect the transition observed.

Can we conclude, then, that the railroad was not important for the American iron industry? The answer is no. This discussion has been concerned with changes in the production of pig iron and has discussed that part of the iron industry that made this material. Rails, however, were made from wrought iron, and the effects of rail production were to be found in the branch of the industry that made wrought iron (and later steel) from pig iron rather than in the branch that made pig iron from ore. The proportion of wrought iron produced used for rails was much higher than the proportion of pig iron so used, and the production of rails had a direct impact on the activities of rolling mills—in contrast to the indirect effect that has been claimed for blast furnaces. The production of rails led to the growth of integrated iron works and the use of the three-high mill in the antebellum era and to the exploitation of the Bessemer process in the years following the Civil War. The use of the Bessemer process was an important change for the iron industry: among its effects were several important changes in the production of pig iron after 1870 that may reasonably be attributed to the demand for rails.<sup>18</sup>

<sup>18</sup> "Hard driving" may be included in this category. See Temin, Part II.

## ANTEBELLUM INTERREGIONAL TRADE RECONSIDERED\*

By ALBERT FISHLOW  
*University of California, Berkeley*

Like Gaul, the antebellum United States was divided into three regions: North, South, and West. Unlike Caesar, however, economic historians have been more concerned with the opportunities thus afforded for regional specialization than the varying warlike characteristics of the populace. From their investigations has arisen one of the abiding generalizations of the structure of the pre-Civil War economy: an industrial North, an agricultural West, and a staple South—all extensively interdependent. This view was summarized by Louis Schmidt in 1939 and has been again brought into prominence by Douglass C. North just recently.

The rise of internal commerce after 1815 made possible a territorial division of labor between the three great sections of the Union—the West, the South, and the East. . . . The South was thereby enabled to devote itself in particular to the production of a few plantation staples constituting a large and growing surplus for the foreign markets and depending on the West for a large part of its food supply and in the East for the bulk of its manufactured goods and very largely for the conduct of its commerce and banking. . . . The West became a surplus grain- and livestock-producing kingdom, supplying the growing deficits of the South and the East.<sup>1</sup>

It is this position I wish to reconsider in this paper. Specifically, I shall hold that the trade between the West and South was always of limited importance to both regions: the South was neither a major market for western produce nor in dire need of imported foodstuffs. Thus to paraphrase (and negate) the comment of Guy Callendar more than half a century ago, the commerce between different agricultural communities in America has played a less important role in our economic history than has recently been argued. Rather, the rapidly growing commerce between East and West played that significant role. In the second part of the paper, with the assistance of a set of interregional trade estimates, I shall show that there was indeed a relative expansion both in regional exports and in their domestic consumption. Interregional exchange was a prominent feature of American antebellum development, but not as a result of interdependence among all regions.

\*I wish to express my appreciation to the National Bureau of Economic Research for assistance in preparing this article during my stay there. A fuller discussion of some of these points may be found in Chap. VII of my *Railroads and the Transformation of the Ante Bellum Economy* (forthcoming). I hope to amplify the contents of the technical appendix and to publish them with further analysis in the near future.

<sup>1</sup>Louis B. Schmidt, "Interval Commerce and the Development of a National Economy Before 1860," *J.P.E.*, 1939, p. 811. Cited by Douglass C. North, *The Economic Growth of the United States, 1790 to 1860* (Prentice-Hall, 1961), p. 103.



*The Trade Between West and South*

For much of the period from 1815 to 1860 New Orleans was the great shipping point for the produce of the interior. The growing volume of receipts at New Orleans, far overshadowing the value of tonnage from the western states arriving at tidewater by the Erie Canal, testifies to that. Unfortunately, however, the produce finding its way to the Crescent City included much beside western exports, as many have recognized in passing. I wonder whether this limitation has been sufficiently appreciated. As early as the 1820's, receipts of cotton, tobacco, sugar and molasses amount to more than half the total value of imports; by the 1850's, these southern commodities made up some three-fourths of the total.<sup>2</sup> For the time when New Orleans was a depository for western products almost exclusively we must go back before the spread of the cotton culture to the western South in the latter 1810's. Thereafter, the trade between the West and the South is most certainly not "recorded in receipts at New Orleans."

This is not to deny that western products did follow the winding course of the Mississippi to New Orleans in significant amounts for much of the period. As late as 1849, it has been estimated that some 40 percent of the western exports of corn, a third of the flour, three-fourths of the salt meat, and two-thirds of the whiskey all were shipped to that southern gateway.<sup>3</sup> Very little was retained for consumption within the South, however. Most was shipped on to northern cities or foreign ports. Table 1 shows the extent of this re-export. Not until the very end of the period were more than half of the principal imports of foodstuffs consumed within the South, including shipments to other southern cities. By then the failure of receipts to keep pace with the rapidly growing export potential of the West meant that consumption of western imports was not notably above the levels established a decade earlier.

Although Table 1 does not go back before 1842, there is no reason to believe that the trend described by it was broken then. We would expect New Orleans to be even more important for its re-export functions the further back we go in time because there was no access to the Erie Canal from the already relatively densely settled Ohio Valley until 1833. There are indications suggesting that this was so. Information upon total receipts at New Orleans and total exports, foreign and coastwise, is available for two earlier years, 1837 and 1833. In the former the exports actually exceeded receipts, while in the latter the

<sup>2</sup>For the 1850's, U.S. Treasury Department, Bureau of Statistics, *Report on Internal Commerce for 1887* (Washington, 1888), p. 209. For the earlier period I have estimated the value of cotton, tobacco, sugar, and molasses receipts and compared these with total receipts given in the *Report*, p. 191 (see Table VII-4, *Railroads*).

<sup>3</sup>See Table 2 *infra*.

TABLE 1  
RECEIPTS AND RE-EXPORTS OF WESTERN FOODSTUFFS AT NEW ORLEANS

Year Ending August 31	Flour (bbls.)	Corn (sacks)	Pork (bbls.)	Bacon (hhds.)	Beef (bbls.)	Lard (kegs)	Whiskey (bbls.)	Value (\$'000 <sup>a</sup> )
<i>1842-45*</i>								
Receipts†.....	491,836	490,169	333,232	30,856	36,023	654,063	81,537	\$ 8,275
Exports†.....	245,542	189,573	244,115	8,012	19,835	575,974	7,274	4,823
Consumption.....	246,294	300,596	89,117	22,844	16,188	78,089	74,263	\$ 3,452
Ratio of consumption to receipts.....	.50	.61	.27	.74	.45	.12	.91	.42
<i>1846-49</i>								
Receipts†.....	1,043,949	1,887,984	507,219	57,760	78,393	1,204,501	126,005	\$20,824
Exports†.....	726,399	1,446,457	308,492	25,303	49,417	1,064,975	15,948	13,830
Consumption.....	317,550	441,527	198,727	32,457	28,976	139,526	110,057	\$ 6,994
Ratio of consumption to receipts.....	.30	.23	.39	.56	.37	.12	.87	.34
<i>1850-53</i>								
Receipts†.....	817,244	1,306,799	441,235	87,378	69,446	1,005,985	140,090	\$22,211
Exports†.....	304,836	496,277	234,578	30,232	49,624	910,169	8,970	10,803
Consumption.....	512,408	810,522	206,657	57,146	19,822	95,816	131,120	\$11,408
Ratio of consumption to receipts.....	.63	.62	.47	.65	.29	.10	.94	.51
<i>1854-57</i>								
Receipts†.....	989,735	1,588,001	325,243	67,658	48,433	702,801	141,424	\$26,300
Exports†.....	529,863	820,267	139,447	16,127	28,340	724,726	7,179	13,053
Consumption.....	459,872	767,734	185,796	51,531	20,093	-21,925	134,245	\$13,247
Ratio of consumption to receipts.....	.46	.48	.57	.76	.41	-.03	.95	.50
<i>1858-61</i>								
Receipts†.....	1,149,695	1,820,616	275,246	63,910	42,287	448,381	139,129	\$24,984
Exports†.....	425,542	410,004	39,543	4,907	15,051	405,351	4,425	6,873
Consumption.....	724,153	1,410,612	235,703	59,003	27,236	43,030	134,704	\$18,111
Ratio of consumption to receipts.....	.63	.77	.86	.92	.64	.10	.97	.72

SOURCE: *Hunt's Merchants' Magazine*, 1842, pp. 391-92; 1844, pp. 419-21; 1845, pp. 370-72; 1846, pp. 406-09; 1847, pp. 413-14; 1848, pp. 511-16; 1849, pp. 554-56; 1850, pp. 536-37; 1851, pp. 602-05; 1852, pp. 489-92; 1853, pp. 624-29; 1854, pp. 475-77; 1855, pp. 601-04; 1856, p. 474 and New Orleans *Price Current*, Sept. 1, 1856; 1857, pp. 603-07; New Orleans *Price Current*, Sept. 1, 1858, 1859, and Aug. 31, 1861; *DeBow's Rev.*, 1860, p. 521.

\* The year 1843 is not included.

† Calculated in homogeneous physical units by dividing total dollar receipts of product by price of physical unit.

‡ Foreign exports plus coastwise shipments to Boston, New York, Philadelphia, and Baltimore. Unspecified coastwise shipments were credited to southern ports.

difference is only slightly positive. This signifies little consumption at New Orleans, and because coastwise trade with other southern cities was limited until the 1850's, affirms the same for total imports of western produce. In any event, total receipts of western produce at New Orleans were very much smaller before the early 1840's—less than half as much in 1836-40 than 1841-45; so consumption necessarily would have been limited even if all foodstuffs were retained.<sup>4</sup>

The significance of the Mississippi to the West therefore was only secondarily as a route to the South. Rather it was a means of reaching eastern and foreign ports, and especially the former; until the later 1850's, twice as much was re-exported to northern cities as to other nations. Laments for the decline of New Orleans as a site of western receipts did not blame declining southern appetites, but, properly, focused on the rapid construction of rail feeders that narrowed the economic hinterland of New Orleans. Nowhere was the shift more obvious than in the Ohio Valley. The proportion of flour flowing eastward or northward from Cincinnati increased from 3 percent in the early 1850's to 90 percent in 1860; similarly for pork, there was a shift from 7 percent to 42 percent.<sup>5</sup> Table 1 shows the combined effects of these shifting loyalties and the more rapid growth of that part of the West tributary to the Lakes: almost twice as much flour, eight times as much pork and bacon, twice as much lard, and three times as much of both corn and beef were exported from New Orleans in 1846-49 than in 1858-61.

These more refined consumption data permit us now to re-examine the trade in foodstuffs between West and South from the standpoint of its relative importance to each region. For the West this means the relative importance of the southern as opposed to the eastern and foreign markets; for the South it means a comparison of the imports of western commodities with the volume of southern production. Table 2, drawing upon the estimates of total western exports of A. L. Kohlmeier, satisfies the first requirement. It demonstrates convincingly the limited extent of the southern market. In total value, less than a fifth of western products were consumed in the South throughout. Only a rising trend of salt meat consumption kept the record as good as this, and here the border states were responsible. In 1853, Louisville contributed a fifth of the pork and bacon received at New Orleans; in 1860 its share had increased to a third.<sup>6</sup>

<sup>4</sup> *Report on the Internal Commerce for 1887*, pp. 199, 215, 285 ff., 377; Thomas S. Berry, *Western Prices Before 1861* (Harvard Univ. Press, 1943), p. 581. Source difficulties aside, one reason why it is not easy to go back before 1842 is the lack of dollar values and specific prices to enable conversion of receipts and exports to homogeneous units. This means an arduous task of reconciling different scales of measurement, as well as lack of weighted annual prices.

<sup>5</sup> Berry, *Western Prices*, p. 91; Israel Andrews, *Trade and Commerce of the British North American Colonies*, House Executive Document No. 136, 32nd Cong., 2nd Sess., p. 711.

<sup>6</sup> Kohlmeier, *Old Northwest*, pp. 118, 202.

TABLE 2  
THE IMPORTANCE OF THE SOUTHERN MARKET TO THE WEST  
(Percent)

Commodity	1839*	1844	1849	1853	1857	1860	1842*	1844	1849	1853	1857	1860
	Proportion of Western Exports Shipped Via New Orleans						Proportion of Receipts of Western Produce at New Orleans Consumed in the South					
Flour.....	53	30	31	27	34	22	42	50	30	60	41	86
Meat products.....	51	63	50	38	28	24	41	31	34	62	69	95
Corn.....	98	90	39	37	32	19	46	70	21	44	65	91
Whiskey.....	96	95	67	53	48	40	80	95	89	90	93	98
Total foodstuffs.....	49	44	40	31	27	17	37	38	29	52	52	85
	Proportion of Western Exports Consumed in the South						Proportion of Western Exports Re-exported Via New Orleans					
Flour.....	22	16	9	14	14	19	31	15	22	13	20	3
Meat products.....	21	19	17	24	19	23	30	44	33	14	9	1
Corn.....	45	63	8	16	21	17	53	27	31	21	11	2
Whiskey.....	77	90	60	48	45	39	19	5	7	5	3	1
Total foodstuffs.....	18	17	12	16	14	14	31	27	28	15	13	3

\* Years ending circa August 31, but not exactly in the case of eastern exports. Note that the 1842 consumption proportions at New Orleans have been applied to 1839.

SOURCE: Proportion of Western Exports shipped via New Orleans: A. L. Kohlmeier, *The Old Northwest as the Keystone of the Arch of American Federal Union* (Principia Press, 1938), pp. 33, 52-53, 83-85, 116-17, 146-48, 191-93, 248-49. Meat products include livestock (estimated for 1839-53). In certain instances there are small inconsistencies between the text descriptions and the summary chart. These have been decided as best possible; they do not affect the results. Total foodstuffs include wheat, which is not shown separately. The proportion is determined from total values as obtained with western prices. See the Technical Appendix (obtainable from the author) for more detail.

Proportion of New Orleans Receipts Consumed in the South: Table 1. Meat products are the weighted sum of bacon, pork, and beef. Total foodstuffs include lard and whiskey, but exclude wheat. This slight incomparability with panel 1 does not influence the findings.

Proportion of Western Exports Consumed in the South: Panel 1 times panel 2.  
Proportion of Western Exports Re-exported via New Orleans: Panel 1 minus panel 3.

If relatively unimportant to the West, the imports were truly minute compared with the production of foodstuffs within the South itself. The 1842 corn consumption of 241,049 sacks (= 2 bushels each) is far less than 1 percent of the 1839 southern crop of 225 million bushels; the corresponding ratios for the census years 1850 and 1860 increase to be sure, but reach a maximum in the latter year of only .9 percent! For wheat the situation is not greatly dissimilar. Wheat imports (principally in the form of flour reckoned as five bushels per barrel), amounted to 960,000 bushels in 1842, 2,600,000 in 1850, and 4,250,000 in 1860. Output was 25 million bushels in 1839, 20 million in 1849, and 38 million in 1859. At best, due to the poor crop of 1849, imports aggregate 13 percent of the total. Note that although imports increased rapidly between 1850 and 1860, local production increased still more rapidly; so the South was less dependent at the end of the decade than at its beginning. Despite the simultaneous boom in cotton, the share of the South in national wheat output actually increased over this interval. Not surprisingly, the relative importance of meat products is more akin to the self-sufficiency in corn. The 1850 census credits to the southern states a product of almost \$49 million in slaughtered animals—an estimate that must be raised to \$119 million to compensate for understatement. Imports of pork, bacon, and beef aggregated only \$3.6 million in the comparable year 1850. In 1860, the ratio of imports, similarly calculated, climbed to the 5 percent mark.<sup>7</sup>

The independence of the South evinces itself even more clearly when contrasted to the role western imports played in satisfying eastern deficiencies. As early as 1849 the 7.5 million equivalent bushels of wheat retained in the East augmented local production by 20 percent. By 1860 the imports exceeded actual output by 20 percent; that is, imports accounted for 53 percent of consumption. In corn and meat products, the East moved from self-sufficiency in the former and imports of a little more than 7 percent in the latter to deficits of 20 percent in both products. Growing eastern demands furnished the where-withal for western expansion; in turn, that region became dependent upon the abundance brought forth upon the lands across the Alleghenies.<sup>8</sup>

<sup>7</sup> These census data for the South exclude Delaware, Maryland, and Missouri. The relationship between reported value of slaughtered meat products and actual value is taken from the ratio of the national census totals to the aggregates reported in Robert E. Gallman, "Commodity Output, 1839-1899," *Trends in the American Economy in the Nineteenth Century* (Princeton Univ. Press, 1960), p. 46. Although the comparisons here have been made in terms of single years, the reader can quickly satisfy himself from the annual averages of Table 1 that the use of a longer interval near the census dates reinforces the conclusions if it affects them at all.

<sup>8</sup> Eastern consumption was calculated as the sum of direct exports to the East as estimated by Kohlmeier, plus re-export to northern cities from New Orleans, minus shipments abroad from eastern ports. For 1860 these last could be obtained directly from the Report on Commerce and Navigation for that year; for 1849 they were estimated as national exports less known exports from New Orleans.

One possibility remains to refute these contentions: that would be a substantial flow of western produce directly to southern sites, and bypassing New Orleans. Such a circumstance is doubtful. The shipments from Cincinnati to down river ports other than New Orleans do not appear to be of any significance. During the years 1848-50, less than 1 percent of the provisions shipped from that city to New Orleans were destined to these other ports; of corn the ratio is 10 percent, but Cincinnati was not the leading southern forwarder; of flour, another growing export of St. Louis, the proportion is about 20 percent on average. At the other side of the transaction, receipts at different southern cities appear limited. Shipments up the Cumberland and Tennessee rivers "supplied a local market in Western Kentucky and were comparatively limited in quantity." The shipment of 297,119 pounds of meat from Vicksburg to Jackson in 1850 is no more than .1 percent of the receipts at New Orleans in the same year. Distribution by rail from Mississippi river points to the interior South likewise is doubtful. The Memphis and Charleston Railroad carried only 4,000 tons of all varieties eastward to the latter city in 1860—too small to count.<sup>9</sup>

Routes to the South other than the Mississippi and the limited channels enumerated above were virtually nonexistent before the war. There was no through connection at all between western and southern railroads and the tortuousness of many transshipments would soon cause exorbitant expense. The underlying economics also make shipment via New Orleans far more probable. Not only was that city a major distribution point, but the limited back haulage would make for low rates upstream. Logic seems to have its counterpart in fact. The *Internal Commerce Report* for 1887 remarks: "There was no trade between the Western cities and the Southern plantations, very little even with the towns; it all paid tribute to New Orleans. . . . Of these shipments upstream over 75 percent. . . were articles which had previously been sent downstream."<sup>10</sup> If the border states of Kentucky and Tennessee be regarded as part of the West rather than the South, some further trade undoubtedly occurred that is not reckoned in here, particularly in livestock. But it is hardly clear that such a treatment is more appropriate: in their commitment to slavery, size of farms, ethnic character of population, and, indeed, in the case of Tennessee its considerable production of cotton, these states were part of the South. Note again, therefore, that our trade figures from the West

<sup>9</sup> *Hunt's Merchants' Magazine*, XXIII (1850), p. 542; Kohlmeier, *The Old Northwest*, p. 202; John H. Moore, *Agriculture in Ante Bellum Mississippi* (Bookman Associates, 1958), p. 111; *American Railroad Journal*, 1860, pp. 840-41 (the through receipts of the Memphis and Charleston were converted to tonnage on the assumption of a 3-4 cent ton-mile rate and a length of road of 271 miles).

<sup>10</sup> P. 205.

actually overstate consumption of western foodstuffs because they include the downriver shipments from Louisville.

Despite this, the clear picture that has emerged is one of tenuous linkage between the two regions. To the West, the South was a minor matter for its own demands. And as early as 1839, more western products were shipped directly eastward than re-exported via New Orleans. Table 2 records the re-export trade in its fourth panel. After 1849 it dwindles rapidly and becomes insignificant by 1860. By then New Orleans was far from the central pivot it had been in the 1810's. On the other side, the southern states were far from dependent upon the agricultural largess of the West for their needs. The greater than average per capita production of corn, peas, and beans in the South supports this observation; so, too, do the larger cattle and swine inventories. The southern social structure, with its large numbers of landowners with few slaves or none at all, also may be invoked. It is suggestive of an economic organization with both widespread self-sufficiency and local sale of foodstuffs to nearby plantations. In conclusion, one may contrast the strength of this case for little southern consumption of western produce with the virtual absence of affirmative evidence for the conventional wisdom of a close interconnection.<sup>11</sup>

### *Antebellum Trade More Generally*

Another perspective from which to view this West-South trade is the context of the other interregional flows. I have set out some estimates of these for various years from 1839 through 1860 in Table 3. Their derivation is described fully in a technical appendix available from the author upon request. Briefly, I may indicate here the nature of these estimates. The trade flows to the West and from West to East directly are obtained by valuing Kohlmeier's shipments of specific commodities, and adjusting for undercoverage; eastern consumption of western products counts in imports via New Orleans and subtracts foreign exports. The flow from West to South is made up of all receipts at New Orleans less imports of cotton, sugar, molasses, and tobacco with consumption determined by the proportion of the selected products of Table 1 retained in the South.<sup>12</sup> The flow from North to South is derived by subtracting the trade from the West to the South from

<sup>11</sup> There has never been unanimous acceptance of substantial interdependence between West and South. Among the prominent dissenters is Isaac Lippincott in his excellent, but neglected, "Internal Trade of the United States, 1700-1860," *Washington University Studies*, IV, Part II, No. 1 (1916). See too the review of North's *Growth of the United States* by Richard Easterlin in the *J. of Econ. Hist.*, 1962, p. 125. George Rogers Taylor, in *The Transportation Revolution* (Rinehart, 1951), comments on the role of New Orleans as a forwarder and the decline of the function, but does not discuss southern consumption explicitly.

<sup>12</sup> This yields an upper bound for southern receipts from the West since residual receipts include hemp from Kentucky, etc. A comparison with the same method used for the East, and the same escalation factor for undercoverage, gives smaller imports in every year. The two estimates are quite close absolutely, however. See the Appendix.

TABLE 3  
INTERREGIONAL MERCHANDISE TRADE FLOWS\*  
(Millions of Current Dollars)

YEAR	ORIGINATING REGION	RECEIVING REGION		
		North	West	South
1839	North.....	—	19.7	85.6
	West.....	11.8	—	14.9
		7.1	—	5.5
	South.....	39.7	6.3	—
		15.1	2.2	—
1844	North.....	—	25.2	73.2
	West.....	20.1	—	19.9
		14.5	—	7.6
	South.....	32.4	6.2	—
		11.6	3.5	—
1849	North.....	—	41.4	80.0
	West.....	36.8	—	36.1
		24.2	—	10.5
	South.....	32.0	8.1	—
		18.9	4.8	—
1853	North.....	—	94.5	147.1
	West.....	63.2	—	36.9
		47.9	—	19.2
	South.....	61.9	17.2	—
		33.0	11.9	—
1857	North.....	—	163.1	165.7
	West.....	96.9	—	49.1
		45.8	—	25.5
	South.....	71.1	13.2	—
		38.0	5.5	—
1860	North.....	—	164.3	213.8
	West.....	146.5	—	42.8
		107.6	—	36.4
	South.....	69.4	20.3	—
		44.6	13.6	—

\* Uppermost figure refers to gross flow whether for consumption or re-export; bottom figure is estimated consumption. In the case of the South-North entry this is limited to northern purchases of southern cotton, sugar, and molasses.

SOURCE: See Appendix (obtainable from the author), Tables A-10, A-11.



estimated total southern import capacity; the South-North exchange is also a residual.

One important limitation of these data should be noted. The flows for the most part exclude full distribution mark-ups since they are valued at the wholesale prices of the region of origin. This procedure has been adopted since the regional allocation of expenditures for freight, insurance, banking, and other services is largely unknown. A measure of the maximum understatement of commerce is given by the extent of interregional price disparity. Sample calculations for both 1839 and 1860 for the West-East trade point to a weighted average differential of 20 percent. Similar calculations for the South-West exchange show a somewhat smaller variation of perhaps 15 percent, depending upon whether southern or western prices are used to value western imports. The trade data of Table 3 are too small by the proportion of this differential accounted for by extraregional payments.

Despite this restriction and the uncertainties in the attempt—due in part to the unsatisfactory nature of some of Kohlmeier's estimates, in part to the variation in quoted prices even within the same region, and in part to the necessity of valuing "merchandise" by the ton—these figures do suggest an order of magnitude that is probably closer to the truth than the variety of other similar figures that are current.<sup>13</sup> Kohlmeier's estimate for western trade in 1844, as quoted from the *Congressional Globe*, is \$120 million for imports and \$115 million for exports. Yet in that same year the Erie Canal valuation for western shipments, which we have used in our derivations, were, respectively, \$14.8 million and \$15.9 million, and that artery carried something like half the exports and even more of the imports, according to Kohlmeier's own estimates. The East-West figures also go beyond the calculations contained in *Statistics of Foreign and Domestic Commerce* wherein both through and way tonnage—much of the latter consisting of local coal destined for eastern seaports—are valued at generous prices to reach what appears to be a rather high value. Similarly, these flows between North and South are an improvement over Kettell's unfounded conjectures of 1860.<sup>14</sup> His total imports into the South are as great as \$462 million whereas the maximum value of cotton, tobacco, rice, and naval stores exports before the Civil War was never

<sup>13</sup> One test of the data is the apparent sense they make with regard to capital inflows to the West (for this region since both imports and exports are derived independently the comparison is meaningful). Table 3 shows the West as a capital importer in 1853 and 1857, as an exporter in 1844 and 1849, and in approximate balance in 1839 and 1860. The first set of years coincide with rapid western expansion, the 1840's with less frenetic extension onto new lands, less railroad and canal construction, etc.

<sup>14</sup> Kohlmeier, *Old Northwest*, pp. 56-57; *Statistics of Foreign and Domestic Commerce*, Senate Exec. Doc. No. 55, 38th Cong., 1st Sess., pp. 129, 181; Thomas P. Kettell, *Southern Wealth and Northern Profits* (New York, 1860), p. 75.

more than \$275 million, against which must be reckoned direct foreign imports of some \$35 million and imports from the West of almost equivalent size.<sup>18</sup>

The data of Table 3 fix the West-South traffic as one of limited scope. Whether we use gross flows or net consumption, the trade always ranks at the lower end of the spectrum. In its meager rate of growth it stands in sharp contrast to the increasing exchange between West and North. Where consumption of western produce by North and South was at approximate parity in 1839, the former was absorbing three times as much by 1860. Gross flows diverge most sharply in the decade of the 1850's as the extension of direct and more efficient East-West transport routes drew the commerce that had once been transshipped from New Orleans. The reciprocal flows inward to the West show the same asymmetrical development. From the earliest, high-valued merchandise was able to bear the cost of transportation and entered from the East, leaving for the South the distribution of assorted groceries and locally produced sugar and molasses. Thus in 1839 the West already depended more heavily upon the East than the South for its imports. By 1860 the advantage had grown enormously; almost ten times as much of western purchases came from the East than upriver. However one reads the record, the tale is straightforward: an initially narrow commerce between the West and the South that failed to keep pace with the rapid expansion in western agricultural bounty and which was supplanted by a total exchange between East and West that ranked first by 1860.

This is not to gainsay the continuing trade between North and South. Over most of the antebellum period the coastwise trade was the most important artery of interregional commerce. Certainly until the mid-fifties the southern market was the largest that faced the North. Yet it is one about which we know perhaps least. Often it is assumed that the South possessed a higher propensity to consume foreign imports, obviously blunting the impact of its demands upon domestic development of manufactures. Yet the sheer size of the market must have given it a key role in American industrialization prior to the 1850's. Philadelphia in particular among the northern

<sup>18</sup> The derivation of these sums is shown in the calculations in the appendix. My imports are understated since exports are valued f.o.b. and hence exclude some of the revenues. If these additional receipts bought imports of services, my import capacity would approximate merchandise imports. These still fall far short of Kettell. I would point out in this connection that such services neither are as large nor was the South as dependent upon the North as is sometime suggested. Commission rates for cotton were low and ocean transport costs a small proportion of value. On the second point, little of the cotton crop was re-exported through northern ports (5-8 percent in the late 1850's), many of the charges for pressing, storage, drayage, etc., were local, and there was as rapid growth of southern banking in the 1850's as nationally, measured by loans and discounts outstanding.

cities seems to have cultivated that territory for its rapidly increasing output of machinery.

Table 3 does more than garb these specific flows with greater statistical precision. It enables us for the first time to examine in quantitative terms the role of interregional trade in antebellum development. Adherents to the tenet that it is the domestic market that counts can take heart from the results. Even if we suppose northern distribution of imported merchandise accounts for a fourth of its shipments to other regions, domestic consumption of the produce of other regions exceeds foreign exports in every year. More significant still is the obvious relative growth in the domestic market during a period of accelerated income advance. Exports to other regions increase from \$109 million in 1839 to \$480 million in 1860; exports to other countries from \$102 million to \$316 million.<sup>16</sup> The regional variation is of interest. The North always fared better in exports to other regions than abroad; conversely the South depended more upon old England than New England for sales of its great staple. The West moves from a southern to a northern stance: in 1839 foreign countries afford a slightly better market for its foodstuffs than the still predominantly agricultural East and plantation South; by 1860, domestic consumption exceeds exports by a substantial margin.<sup>17</sup>

Export base proponents also can find much to their taste. The sum of the exports to other regions and abroad increases in importance between 1839 and 1860. Relative to gross national product in the former year they are 13 percent; in the latter 19 percent. Over this period their potential influence was therefore considerably enhanced. Again, the specific regional patterns deserve mention. Southern exports increased only slightly, from 23 percent to 29 percent of income. The West's dependence upon an external market radically altered from less than 13 percent to 23 percent of its product. The northern export percentage moved intermediately from 10 to 15. The region displaying the greatest increase in income and population, the West,

<sup>16</sup> The total consumption of domestic produce of other regions is the sum of the bottom entries of Table 3 plus three-fourths the flow of eastern merchandise to other regions. Western imports of salt and iron from the East are treated as of domestic origin while coffee and sugar are treated entirely as foreign; all trade from North to South is treated as merchandise. Since the ratio of imports of manufactures to census value of eastern domestic production (minus the two important processing industries of sawing and milling) is less than .2 in 1849 and 1859, the adjustment is not likely to overstate the domestic market. For exports and imports by class, see North, *Economic Growth*, pp. 284, 288, and for detail on western imports from the East see the Appendix (obtainable from the author).

<sup>17</sup> These comparisons consist of the sum of consumption of western products by North and South as given in Table 3 and the national exports of breadstuffs and provisions, amounting in millions of dollars as follows: 1839, \$14.1; 1844, \$18.0; 1849, \$38.2; 1853, \$33.0; 1857, \$74.7; 1860, \$45.3.

also had the stimulus of the most rapidly growing external market, exactly as export base theory would predict.<sup>18</sup>

### *Final Comment*

Full exploration of these intriguing suggestions is beyond the scope of this paper. But before the analysis of the role of interregional trade in American antebellum development can press home, further advances in two directions will be required. Refinement and extension of the basic trade estimates is an essential step; a promising possibility is a complementary approach from the side of regional consumption requirements. Equally urgent is modification of the export base theory to render it more amenable to empirical test. At present it does not completely satisfy this criterion.

Such a restatement can draw upon some of the preliminary findings reached here. The transition of the West from reliance on a foreign to a domestic market, in contrast to continued southern dependence on overseas markets, suggests that differences in destination of exports may play a greater role than has been recognized. Exports to another region may be more stable and less subject to competition than sales on a world market; regional interdependence may also bring with it a greater reciprocal inflow of labor, capital, and institutional influences. Whatever the merit of this speculation, the approach must certainly come to grips with our central result: the small, and lessening, linkage between the West and the South. Interregional interdependence there was before the Civil War, but increasingly exclusive of the South, a circumstance not unrelated, one might add, to the decision for Confederacy by the South, and that for Union by the West.

<sup>18</sup> The current dollar gross national product estimates are those of Robert E. Gallman in "Gross National Product in the United States, 1834-1909," to be published in Volume 29 of *Studies in Income and Wealth*. They were converted to regional values for 1840 and 1860 by the use of Richard A. Easterlin's regional relatives for the two dates as given in Seymour Harris, ed., *American Economic History* (McGraw-Hill, 1961), p. 528. Western exports are approximated by the sum of domestic consumption in South and North plus foreign exports of breadstuffs and provisions; southern exports are made up of sales of cotton, tobacco, rice, sugar, naval stores, and molasses (see Appendix, Table A-7); northern exports to other regions are estimated as in fn. 17 to which are added national exports of manufactures and semimanufactures. For both North and West the results are not exact, but they should be close to the correct magnitudes.

## CANALS AND DEVELOPMENT: A DISCUSSION OF THE ISSUES

By ROGER L. RANSOM  
*University of Virginia*

It is widely recognized that one of the important factors promoting economic growth in the United States during the first half of the nineteenth century was an increased degree of regional specialization according to the comparative advantage of each region. The degree of economic specialization in a region is limited by the extent of the market; and one of the principal constraints on the size of markets has been the cost of transportation between regions. Economic historians have therefore focused considerable attention on the development of an efficient transportation system in the United States. Traditionally the development of the railroad has been hailed as the most important impetus to expansion of markets in America; however, recent research has tended to throw some doubt on this conclusion. The existence of an extensive system of canals prior to railroad construction has caused a revived interest in the role of canals in economic growth. This paper is concerned with the issues which are raised in attempting to assess the contribution of the canals to the economic development of the United States.

### I

At the time of the Louisiana Purchase, the Appalachian Mountains divided the United States into two distinct geographical regions. Water routes connected the principal primary markets within each region; however, communications between the East and West were hampered by the very high cost of transportation around or through the mountains. A successful breach in this barrier was not made until 1825, when the completion of the Erie Canal provided an all-water route from the Great Lakes to the Atlantic Seaboard. The impact which the Erie had on the country was enormous. It provided dramatic proof that a canal to the West was both technically feasible and economically profitable. Even before the Erie was opened for its full length, its interest costs were being paid from revenues on sections already opened.

The Erie provided a serious challenge to other sections of the country. To the eastern cities it threatened to monopolize the lucrative western trade in the hands of New York. To the West it provided a challenge to expand their markets by supplementing the Erie with

other canal routes. The result was a period of canal investment continuing with varying gusto for forty years. Recent research by Jerome Cranmer and Harvey Segal has shown that this investment occurred in three distinct cycles.<sup>1</sup> The initial cycle lasted from 1816 to 1834 and involved the construction of 2,188 miles of canal at a cost of \$58.6 million. A considerable amount of the investment in this first cycle was in direct response to the Erie's success. The second cycle, from 1834 to 1844, was largely concerned with the construction of canals designed to service the major trunk lines constructed in the preceding period. Investment in this cycle totaled \$72.2 million for 1,172 miles of canal. The final phase, lasting until the Civil War, was for the most part an "echo phenomenon" from earlier cycles.<sup>2</sup>

By 1860, therefore, about 4,250 miles of canal had been constructed, representing a total investment of almost \$190 million. Yet canals in 1860 were only one part of the overall transportation network of the United States. Even before the canal era the economy had been provided with an extensive system of natural waterways. The Mississippi River and its tributaries provided navigable waterways in almost all of the western states, and the Great Lakes provided northern communications between western areas. On the Atlantic Coast were a large number of good harbors, and many rivers stretching to the interior. The introduction of the steamboat in 1816 made these river routes into valuable avenues of transportation throughout the antebellum period. By 1850 a second form of transportation had arisen to challenge the canal: the railroad. In 1860 there were 30,000 miles of railroad track in the United States, two-thirds of which had been constructed in the last decade.

It is, in fact, illusory to speak of a "system" of canals. The network of canals which had emerged by 1860 was not a cohesive, well-planned system constructed with a common aim. Portions of it represented attempts to complement the existing facilities, while portions of it represented attempts to compete with existing (or potential) facilities. These efforts were extended over a long period, and there were errors of judgment and much duplication of effort. These errors are reflected in the great diversity of canal performances: there were spectacular successes and there were spectacular failures. The investment in canals should not be evaluated as comprising a system. Each canal project should be evaluated on its individual merit, in terms of the alternatives existing at the time of its construction. The issue is not the suc-

<sup>1</sup> Jerome Cranmer, "Canal Investment, 1816-1860," in *Trends in the American Economy in the Nineteenth Century* (Princeton, 1960); and Harvey Segal, "Cycles of Canal Construction," in Goodrich (ed.), *Canals and American Economic Development* (Columbia Press, 1961).

<sup>2</sup> The term is Segal's.

cess or failure of the canal system; it is the extent to which various individual canals represented a wasteful or useful allocation of resources in a growing economy.

Even a cursory glance at the records of canals in the United States shows that many of them—including some of the major interregional canals—were not profitable to their operators. Economists are quick to point out that financial returns alone constitute an inadequate measure of the contribution of these canals to the economy. There are “indirect effects,” which if taken into account would show that many more canals were profitable from the social standpoint. Much has been written concerning these indirect effects; yet little attempt has been made to measure them. Indeed, it is often argued that such benefits cannot be quantitatively estimated. If this is the case, then we can never accurately assess the contribution of the canals to growth. Clearly the magnitude of these external effects lies at the very heart of the issue of whether the return from the canals was sufficient to justify the capital investment required to construct them. I am less pessimistic than some of my predecessors. I think that meaningful estimates of these returns are possible, and I propose to discuss several techniques of doing so. First, however, I want to indicate what I consider to be the major issues in the discussion of canals and growth.

In evaluating the contribution of canals to growth, historians have tended to combine two essentially different sets of issues. The first involves a problem of technological obsolescence. The technological superiority of the railroad meant that the returns to the canals were often reduced to unprofitable levels before the initial capital investment could be repaid. Given this premature obsolescence, were many canals able to provide sufficient returns in their shortened life to justify their construction? This issue deals with the actual contribution of the canals. It involves an evaluation of the importance of the water transportation to the economy and the part which the canals represented in the system of internal waterways. I argue that it is necessary to consider each of the canals as an individual investment in the context of the overall system of transportation—water and rail. We must then determine which canals were socially profitable. Were any canals, although unprofitable themselves, necessary to make other canals successful? Which canals were not successful? We must not simply consider which canals succeeded or failed; we must try to estimate the magnitude of the social gain (or loss) from the canals. From these individual evaluations, a picture of the role played by canals in the development of the economy will emerge.

The extent to which canals were in fact the victims of technological obsolescence will also be indicated. It is possible that many canals

would have failed even without the railroads. This brings up a second issue: the extent to which the canals, in the absence of railroad competition, would have contributed to growth. This issue is attempting to measure the potential return from the canals. How large would the returns from the successful canals have been with no rail competition? What canals which were failures because of the railroads would be successful without them? Which canals would have failed in any event? Were there any canals which might have profitably been constructed? These questions represent more than idle speculation on the probable course of events. They indicate the extent to which canals could have provided the needs of the economy as it grew. The answer would be useful in assessing the contribution of the railroad to growth. To properly assess the role of the railroad, we must be able to indicate the alternative situation which would have existed. These second questions provide this answer.

The issues are, I think, substantial ones. I cannot hope to answer them here, and I shall therefore direct my remarks to several specific areas of the canal discussion.

## II

Whichever issue you choose to investigate, the solution will depend on being able to estimate the external returns from the canals. While the problem of measuring these benefits is a difficult one, it is not insurmountable. One can consider the gains from more efficient transportation from the standpoint of alternative cost. Such an approach has been suggested by Robert Fogel in his investigation of the railroads and American growth.<sup>3</sup> If the economy did not have the canal, then the national product would fall by an amount equal to the added real cost of shipping the same bundle of goods to the same destinations via the least cost alternative to the canal. Given data on shipments and costs, a linear programming model can be constructed to indicate the least cost solution to this transportation problem. It may not be the solution actually chosen by the economy; however, it will be the best possible one. This model can be solved first with the canal included as a possible means of transportation; then recomputed with the canal excluded as a possible route. The difference in cost between these two solutions represents the amount which the canal "saved" the economy. This saving can be interpreted as the external benefits from the canal.

Fogel terms this figure the " $\alpha$  estimate of social saving." It represents the upper bound to the magnitude of the fall in income. If in

<sup>3</sup> Robert Fogel, "A Quantitative Approach to the Study of Railroads in American Economic Growth," *J. of Econ. Hist.*, June, 1962.



fact the canal had not been built, the economy would adjust its allocation of factors in response to the new pattern of transportation costs. Such an adjustment would lessen the fall in income, and the true saving would be less than the  $\alpha$  estimate. With a small value of social saving (such as Fogel derives in his study of railroads and interregional trade), this upward bias would be no problem. When the alternative costs of shipping are very high relative to the costs via the canal, the overstatement of benefits is likely to be more serious. Failure to allow for adjustment to the new costs of transportation might result in the situation where goods are being shipped to market even though the transportation costs are greater than the value of the goods at market. The upper bound estimate may then be so far above the true value of the benefits that it fails to provide an adequate guide to evaluate the canal returns.

A new estimate (which Fogel terms  $\beta$ ) is required.<sup>4</sup> Without the canal, there would be some "boundary" of feasible production beyond which commercial agriculture would cease to be profitable. The social saving from the canal would then be the  $\alpha$  estimate for goods shipped from within this boundary, plus the loss in income to the economy resulting from the inability to use the land which is outside the area of feasible production. In his forthcoming essay, Fogel shows that if data on the pattern of shipments, the costs of transportation, and the value of land in the region are available, then the  $\beta$  estimate can be computed. By allowing for a partial adjustment to the non-canal situation (i.e., not considering goods shipped from outside the area of feasible production in the computation of the  $\alpha$  estimate), the  $\beta$  estimate partially removes the source of extreme upward bias in the  $\alpha$  estimate. The bias of the  $\beta$  estimate is uncertain.

Land value data have been used for a long time as an index of the external benefits resulting from transportation improvements. If transportation costs fall, the price received at the farm increases. This increases the return to the land, and the price of land rises. However, forces other than transportation costs can affect the value of land. To use a change in land values as an index to measure the external effects of the canal, one must be able to account for any other forces which might cause land values to change. Even without data on land values, which are necessary for the  $\beta$  estimate, the increase in return to the land resulting from the canal might be estimated. The shipments along the canal will reveal that point which represents the farthest distance over which a good has been shipped to a primary market. This point

<sup>4</sup> See Robert Fogel, "The Intra-regional Distribution of Agricultural Products" (preliminary draft of an unpublished paper). I am indebted to Mr. Fogel for allowing me to read this manuscript prior to publication.

will be on the boundary of feasible production existing with the canal. If the rent here is assumed to be zero, then the cost of shipping the good to market via the canal is the maximum transportation cost which that commodity can bear. Producers closer to the market (in terms of transportation cost) would receive economic rent on their shipments. In the event that the producer is beyond the area of feasible production which would exist without the canal, we can infer that all of this rent is attributable to the canal, since without the canal his rent would be zero. If, on the other hand, the producer is within the area of feasible production existing without the canal, his rent would be positive even if the canal were not built. In this instance, the return under the non-canal situation would have to be subtracted from the estimated rent with the canal. The resulting "net rent" would correspond to the  $\alpha$  estimate of social saving when the goods could be shipped to market even without the canal.<sup>5</sup> Although the data requirements for the estimate of inferred rent are less stringent than the requirements for the social saving estimates, the assumptions are more restrictive. Only if the quality of the land in the region is homogeneous will the rent estimate be a true reflection of the benefits contributed by the canal. Variations in the quality of land, which are taken into account through the use of land value data, are not accounted for in the case when rent is made a function of location alone.

Both the  $\beta$  estimate and the rent estimate contain biases whose directions are not always clear. Problems may arise in applying these estimates to aggregate returns, since both are derived from a partial equilibrium analysis. Given a general equilibrium situation, prices of the commodities at the primary markets would change as demand and supply functions changed. Such changes in price would affect the estimate of returns from either the  $\beta$  estimate or the rent estimate. The  $\alpha$  estimate will always provide an upper bound to the value of the benefits; however, what is desired is an estimate which will be as close to the true value of returns as possible, and the  $\alpha$  estimate will in many cases be very wide of the mark.

Insufficient information concerning the exact pattern of shipments may prevent the derivation of any of the above estimates. While data may not be sufficiently accurate for a linear programming model, it may permit us to estimate the total volume of traffic in ton-miles. In such a case, a crude estimate of the  $\alpha$  estimate of social saving can be calculated by multiplying the volume of traffic (ton-miles) by the average saving per ton-mile (\$). Such an estimate is clearly inferior to the

<sup>5</sup>For a more complete discussion of the "rent estimate" see R. L. Ransom, "Government Investment in Canals: A Study of the Ohio Canal, 1825-60" (unpublished Ph.D. thesis, University of Washington).

$\alpha$  estimate itself. All of the upward bias of the sophisticated estimate remains, and an additional upward bias is introduced by the fact that the crude estimate does not even permit the economy to utilize the least cost alternative in the non-canal situation. Only in instances where the social saving is very small would such an estimate provide an approximation to the true return.

The discussion above is intended to indicate that while the problems involved in estimating the indirect effects of a canal (or railroad) are formidable, they are not insoluble. The techniques above yield meaningful economic estimates of the savings from a fall in transportation costs. A major characteristic of all the techniques is their emphasis on the alternative to the canal in determining the size of the indirect effects. The least cost alternative is not something which will remain constant over the life of the canal. The development of a more intensive system of transportation will create new opportunities to reduce the alternative cost to the canal, and hence affect the indirect benefits.

### III

One of the few attempts to estimate quantitatively the returns from the canals in the United States has been made by Harvey Segal.<sup>6</sup> He argues that canals were instrumental in opening the West to the markets in the East and allowing the regional specialization to proceed at a rapid pace even before the appearance of the railroad. In the last part of his study he presents an estimate of returns from a sample of canals designed to show that these canals were clearly profitable to the economy. Segal's study is worth careful consideration because I think it points out the problems of confusing the issues I mentioned above.

Segal applies cost-benefit analysis to evaluate the performance of ten "heavily utilized" canals in the decade 1837-46. This period is deliberately chosen to evaluate the canals prior to railroad competition. He estimates the annual volume of traffic on these ten canals (288.1 million ton-miles), multiplies it by the average saving per ton-mile (23 cents), and estimates that the upper bound to the annual benefit from these canals was \$66 million. I have already pointed out that this technique involves a very substantial overstatement of benefits. Segal is aware of this, and so he approaches the problem from the cost side. If the costs can be computed, then the necessary benefits to equate discounted costs and benefits can be computed. An estimated \$128 million was invested in canals prior to 1846. With a period of amortization of fifty years and an interest rate of 8 percent the annual capital costs of the canal system would be \$10.5 million. Considering the net revenue from the canals increases the annual costs to \$12.5 million or 3.6 cents

<sup>6</sup> Harvey Segal, "Canals and Development," in Goodrich, *op. cit.*

per ton-mile.<sup>7</sup> This would be the annual benefit necessary to equate costs and revenue. Surely, Segal concludes, this is a sufficiently small figure to have been readily met by the canals in his sample alone.

But is 3.6 cents per ton-mile an insignificant figure? When Segal changed from his discussion of benefits to his discussion of costs he implicitly changed the question he was asking. When he computed the benefits on the basis of traffic before rail competition, he was concerned with the issue of the potential return from the canals without the railroad. When he chose a period of amortization of fifty years, however, he was including at least thirty-five years of railroad development. An annual saving of 3.6 cents per ton-mile on these canals from 1837 to 1846 is not impossible (although I consider it more improbable than Segal apparently does); but an annual saving of 3.6 cents per ton-mile from 1837 to 1886 is quite improbable. Yet to amortize the canals, it is the latter condition which must be met. If the actual contribution of the canals is being considered—and this appears to be Segal's intention—the costs and benefits must both be measured over the same period. A period of amortization of twenty or twenty-five years appears to be somewhat nearer the actual life of most canals. Considered over twenty-five years the annual benefits would probably decline and the costs would increase. The answer to the question of whether the canals in fact repaid their capital is much more dubious than Segal's treatment of the problem indicates. He has shown that in the absence of railroad competition, the investment in canals prior to 1846 could have been repaid from the earnings of the ten canals in his sample.

Even if this were so, Segal recognizes that the profits from successful canal investments should not be used to justify canals which were unsuccessful. His data on the volume of traffic indicate the extent to which canals were not all equal. Of the total traffic on all ten canals, the Erie Canal alone accounts for almost one-half of the total. A similar picture is conveyed by a statistic presented earlier in the study stating that from 1835 to 1853 the Erie's share of canal traffic from the West to the East rose from about 35 percent to 60 percent. The fact that a single canal so dominated the aggregate statistics leaves room to suspect that aggregation hides as much as it reveals. Although the Erie contributed a substantial portion of the benefits, it represented only a small portion of the total investment in canals. Removing the effects of the Erie from the sample makes the profitability of the remaining nine canals much more doubtful.

<sup>7</sup> In the original publication of his essay, Segal miscalculated the annual costs as \$1.05 million. This was corrected in an errata sheet; however, he did not adjust the conclusions which he drew from the earlier estimate, despite the increased costs indicated by the correction.

## IV

I am not prepared at this moment to offer detailed quantitative estimates of the returns to various canals in the United States during the antebellum period. Segal's data indicate to me that many canals were not profitable. He estimates that \$40 million of canal investments failed. I feel this estimate is quite conservative. To indicate the basis for my skepticism of the proposition that canals were profitable, let me outline the case of the Ohio Canal—one of the more successful of the state canal ventures.<sup>8</sup>

The Ohio Canal was completed in 1833, well before the advent of the railroad in the area. It was carefully planned to connect the Ohio River with the Erie Canal and the eastern markets. Over the interval 1836 to 1850, the annual external benefits from the major agricultural commodities carried on the canal would average about \$1.85 million per year if the crude estimate of social saving is employed. Over the same period, the identical commodities yielded an estimated rent which never exceeded \$270,000 in any year. I would argue that in this instance, where there is no practical alternative to the canal, the rent estimate is a more accurate reflection of the true benefits than the crude estimate of social saving. Including net revenues, the average return from 1836 to 1850 was about \$440,000, or 11 percent return on the invested capital. Had this level been maintained over a longer period, the capital investment would have been repaid. However, it did not continue. In 1851 the railroad was completed to several points on the canal, and by 1855 the major towns on the canal had railroad service. The impact of the competition on the canal returns was sudden. From a level of return of 15.4 percent in 1851 (using the rent estimate), the return to the canal fell to 2.7 percent in 1854-56. The maximum estimate of return in 1854-56, using the  $\alpha$  estimate, was only 5.2 percent. By 1859-60 the maximum estimate of return on the canal was zero; the estimate using rent was negative. If one discounts the net revenues plus estimated rent from the canal to 1860, it failed to repay its capital. The rent estimate is not an upper bound, and it is still possible that the Ohio Canal had a sufficient rate of return to be profitable; however, the point is not obvious.

I think two points emerge from the example of the Ohio Canal. First is the speed with which the railroad reduced the returns of this prosperous canal to an unprofitable level. The Ohio Canal operated profitably for over fifteen years. On canals less established than the Ohio, an equally sudden termination of returns might leave substantial portions of the capital investment unpaid. The second point is the extent

<sup>8</sup> All of the data on the Ohio Canal are from Ransom, *op. cit.*

to which the social saving estimate (and many contemporary judgments based on a similar notion) overstates the returns in situations where the alternative cost of transportation is very high. There can be little doubt that the Ohio Canal was a major influence in shaping the economic development of eastern and central Ohio in the 1830's and the 1840's. But in the absence of the canal, adjustments would have occurred. Production would have centered along other transportation routes. While much of the interior area of eastern Ohio would remain isolated from the market for an additional decade, this does not mean that all of the production in these areas would necessarily have been lost to the economy as a whole. The effect of the canals on a specific region as opposed to their effects on the entire economy must be kept apart. Even when adjustments are taken into account, the benefits from the Ohio Canal might have justified its construction. However, I think the magnitude of benefits would be considerably less than is commonly indicated.

The role of the state in initiating many of the important developmental canal projects has been frequently mentioned in the literature. In terms of traffic and revenues from tolls, the Ohio Canal was far more successful than most state ventures. Indeed, if one considers the record of state canal investments in the United States, it is not an impressive one. In Table 1 below, I have grouped \$102 million of state canal investments according to whether I judge them "probably successful" or "probably unsuccessful." The grouping is based on data regarding traffic and revenues judged in the light of the Ohio Canal experience. Such a grouping is inevitably imperfect; however, I hazard to guess that more detailed research would not be more favorable to the government's record. Of the \$102 million, I find that only about \$16 million was profitably invested in successful canals. This leaves \$86 million of unsuccessful canal investments. This figure should be adjusted for some sections of the unsuccessful canals which might have repaid their investment. While the Pennsylvania Mainline Canal (including the railroad connections), the Miami and Erie Canal, and the Wabash Canal were probably not successful when considered as whole units, some sections which were more heavily used or opened before construction of the entire canal was completed might have been profitable. The total investment represented by such sections would probably not have exceeded \$15 million. Adding this to the group of possible successful canals gives a figure of about \$31 million of profitable investment which is offset by an estimated \$71 million of unprofitable investment. Naturally, some of the \$71 million invested in unsuccessful canals was repaid from earnings on these canals. The extent to which

TABLE 1  
GOVERNMENT INVESTMENTS IN CANALS

Canal	State	Cost (\$000)
I. Canals which were probably successful:		
Erie Canal.....	New York	7,143
Champlain Canal.....	New York	921
Oswego Canal.....	New York	2,512
Ohio Canal.....	Ohio	4,245
Delaware Division Canal.....	Pennsylvania	1,543
Total successful canals.....		16,364
II. Canals which were probably not successful:		
Miami and Erie Canal.....	Ohio	5,920
Wahlhonding Canal.....	Ohio	607
Hocking Canal.....	Ohio	975
Wabash and Erie Canal.....	Ohio	500
Black River Canal.....	New York	3,157
Genesee Valley Canal.....	New York	5,663
Chenango Canal.....	New York	2,316
Mainline Canal*	Pennsylvania	16,473
5 Penn. Lateral Canals†.....	Pennsylvania	15,033
Wabash Canal.....	Indiana	6,325
Whitewater Canal.....	Indiana	1,400
Illinois and Michigan.....	Illinois	6,558
Chesapeake and Ohio‡.....	Maryland	11,071
James and Kanawha§.....	Virginia	10,436
Total unsuccessful canals.....		86,434
Total canal investment.....		102,798

\* Mainline Canal cost includes railroad connections.

† The five canals were: the Susquehanna Division Canal, the French Creek Canal, Beaver Canal, the North Branch Division Canal, and the West Branch Division Canal.

‡ Private company whose stock was largely owned by Maryland, Virginia, and the United States government.

§ The \$5.5 million of stock was purchased by Virginia, Richmond, and Lynchburg.

this is the case would be an important factor in determining the extent of the misallocation of resources.

## V

In summary, I conclude that the role of canals in contributing towards the economic growth of the United States is far from obvious. Clearly some canals, such as the Erie and the "coal canals" of Pennsylvania and New Jersey, were valuable investments which contributed considerable impetus to growth. However, a substantial number of the canals constructed before the war were clearly not profitable—even when external effects are considered. Too many canals were competing with other transportation facilities—canals or railroads—rather than opening new avenues of trade between regions. To be sure, the gains from regional specialization in the United States during this period were substantial. More productive land in the west was brought under cultivation, and the cost of producing an equivalent supply of food in

the east would be substantial. But I think the evidence indicates that the market could have been widened to allow much of this specialization at a cost well below \$190 million. Far too much attention has been paid to the regional benefits rather than the increased efficiency of the economy as a whole.

But my purpose here is not to answer the questions which I raised initially. While I have conveyed my personal skepticism concerning the contribution of the canals to growth, I would be the first to admit that such a judgment remains to be confirmed (or refuted) by further research. I have attempted here to outline the general approach which I feel that research should take. Considerable effort has been expended in recent years to investigate the magnitude of canal investment and the role of the government in promoting canals. Now we need to pursue that research one step further and ask: Was the investment in canals worth the cost?



## DISCUSSION

ROBERT WILLIAM FOGEL: In a note published in the March, 1963, issue of the *American Economic Review* Professor Douglass C. North stated that there was a "revolution" under way in the writing of economic history in the United States. According to Professor North the revolution has been "initiated by a new generation of economic historians who are skeptical of traditional interpretations of U.S. economic history and convinced that a new economic history must be firmly grounded in sound statistical data."<sup>1</sup> I am not sure if Professor North coined the term "new economic history" or was merely the first one to put it into print. In any case the use of the term has become rather widespread in recent months.

The papers presented here were written by three of the youngest and most promising practitioners of the new economic history. I want to discuss their papers as typical products of the work of the new generation to which Professor North referred in order to see if this work is revolutionary in some sense. More particularly, I wish to consider whether or not one observes in the work of the new generation a departure from the writings of the past which is sufficient to justify a title like new economic history.<sup>2</sup>

It has been said that the element which distinguishes the new economic history is an emphasis on the quantitative aspects of economic processes. If all one means by this is that a major effort is made to discover and present numerical information relating to historical processes, then the new economic history is not very new. Many of the outstanding studies of the past abound in the presentation of such data. Gray's *History of Agriculture in the Southern United States*, Clark's *History of Manufactures*, Paxson's "Railroads of the Old Northwest," Bogart's "Early Canal Traffic and Railroad Competition," Hunter's *A Study of the Iron Industry of Pittsburgh Before 1860* are just a few of the works that fall into this category.

Indeed, each of the speakers has relied heavily on the numerical information collected by one or more of the "old" economic historians. Fishlow uses Gray's series on rice and tobacco prices at New Orleans in deriving his estimates of southern import capacity. And his figures on shipments between the West and the East are based primarily on data taken from Kohlmeier's *The Old Northwest*. Price data collected by Hunter play a critical role in Temin's explanation of the slowness of western ironmasters in increasing the production of pig iron made with coke.<sup>3</sup> Ransom uses railroad distances compiled by Paxson in computing his estimates of the unpaid benefits of the Ohio Canal.<sup>4</sup>

There is in this respect, and in others, a clear line of continuity between

<sup>1</sup> "Quantitative Research in American Economic History," *A.E.R.*, Mar., 1963, p. 128.

<sup>2</sup> If discussants were permitted to use titles, these remarks would be called: "A Provisional View of the 'New Economic History.'"

<sup>3</sup> Albert Fishlow, "The Economic Contribution of American Railroads Before the Civil War" (unpublished doctoral dissertation, Harvard Univ., 1963). Peter Temin, "A History of the American Iron and Steel Industry" (unpublished doctoral dissertation, M.I.T., 1964).

<sup>4</sup> Roger Leslie Ransom, "Government Investment in Canals: A Study of the Ohio Canal, 1825-1860" (unpublished doctoral dissertation, Univ. of Washington, 1963).

the old and the new economic history. However, in pursuing their quantitative interests, the old economic historians limited themselves primarily to the presentation, in more or less original form, of data found in standard historical sources. Very little was done to transform the original data in a way that would make it shed light on "rigorously defined concepts of economic analysis."<sup>5</sup> Moreover, the old economic historians limited themselves almost exclusively to measuring things that could be measured directly. Things that could be measured only indirectly were often discussed in qualitative terms but discussions of the quantitative aspects of such entities were rare.

It is in the redirection in measurement that one observes a distinct departure in the work of the new economic historians. Although like their predecessors the new economic historians devote a considerable proportion of their energy to the location of neglected sources of economic data and to various simple classifications of the evidence they find, this is not their primary thrust.<sup>6</sup> They have evolved a series of new preoccupations. One of these is the effort to reconstruct economic measurements which might have existed at some time in the past but which are no longer extant. An interesting attempt at such a reconstruction will be presented at the Econometric Society's panel on economic history. Professor Paul David in a paper entitled, "Economic History through the Looking Glass," will discuss his effort to infer from output, employment, and wage data which have survived the passage of time, missing information on the growth of the stock of capital in Chicago between 1870 and 1893.

The new economic historians are also concerned with the problem of recombining primary data in ways that enable them to obtain measurements that were never before made. Like many of his predecessors Fishlow uses data published in commercial journals on receipts and shipments at New Orleans in order to shed light on the pattern of interregional trade. But what was a final product for other scholars is to Fishlow one input into a more ambitious attempt to arrive at aggregate measures of interregional trade flows in the antebellum era. Similarly, Temin has numerous references to the cost of various fuels and to the prices of various types of pig iron in much the same way as Hunter and Taussig. But Temin goes beyond them in attempting to use such data to compute an index of the average rate of profit earned by the producers of the different types of metals.

In the course of their attempted constructions and reconstructions the new economic historians often come to points at which information on a process or institution essential to the completion of a given statistical structure is largely or completely missing. While this might seem to doom a particular enterprise, it need not. For frequently a measurement which is logically neces-

<sup>5</sup> Simon Kuznets, "Summary of Discussion and Postscript," *J. of Econ. Hist.*, Dec., 1957, p. 553.

<sup>6</sup> However, high-speed computers have opened up important new possibilities even in circumstances where all that is required is a simple arithmetic operation on, or classification of, existing bodies of data. See Lance E. Davis, Jonathan R. T. Hughes, and Stanley Reiter, "Aspects of Quantitative Research in Economic History," *J. of Econ. Hist.*, Dec., 1960, pp. 531-47; and Harvey H. Segal and Matthew Simon, "British Foreign Capital Issues, 1869-1894," *J. of Econ. Hist.*, Dec., 1961, pp. 566-81.

sary for a given computation may be such that any number regardless of how much it departs from reality is permissible and serves to close the logical system required to support a construction or reconstruction. Professor Fishlow is particularly adept in the exploitation of this device. Thus in arriving at an estimate of aggregate shipments from the North to the South, Fishlow has to make some assumption about the share retained in the South of certain products received at New Orleans. He arbitrarily assumes that the retention ratio appropriate to these commodities is the same as the average retention ratio that was observed to have prevailed on another group of commodities. However, even a 40 percent error in Fishlow's assumed ratio would affect his estimate of the volume of North-South trade by only a few percentage points. Not only in Fishlow's work but also in the studies underlying the papers of Ransom and Temin as much attention is devoted to the possible consequences of missing data on their statistical constructions as is devoted to the implications of the evidence that is available.<sup>7</sup>

Still another departure in the approach to measurement characteristic of the new economic history is the emphasis placed on finding methods of measuring that which cannot be measured directly. Considerable attention was devoted to this problem by Alfred H. Conrad and John R. Meyer in their essay on "The Economics of Slavery in the Ante Bellum South"—an essay which has assumed the stature of the first classic of the new economic history.<sup>8</sup> Conrad and Meyer used indirect measurement in evaluating the hypothesis that southern attitudes prevented a widespread internal slave trade. They argued that if social or psychological barriers effectively thwarted trade, one should find approximately the same demographic structure in the slave population in all states. However, census data revealed that the "buying" states of the South had a larger proportion of slaves in the prime working ages than the "selling" states and the selling states had a larger proportion of slaves in the ages under fifteen and over fifty.<sup>9</sup>

Indirect measurement is the foundation of Professor Ransom's evaluation of the social profitability of canals. As Professor Ransom points out, one cannot judge the efficacy of canal construction without arriving at some conclusion concerning the size of the unpaid benefits of these waterways. One possible approach to the determination of the unpaid benefits turns on the estimation of the increase in the rental value of land traversed by a new

<sup>7</sup> An attempt to establish criteria for statistical reconstructions based on fragmentary data is contained in Robert William Fogel, *Railroads and American Economic Growth: Essays in Econometric History* (to be published by the Johns Hopkins Press, fall, 1964), Chap. V. For other views of the potentialities and limitations of synthetic reconstructions see: Conference on Income and Wealth, *Trends in the American Economy in the Nineteenth Century, Studies in Income and Wealth*, Vol. 24 (Princeton Univ. Press, 1960), especially the introduction by William N. Parker, the exchange between Robert E. Gallman and Neal Potter, and the comments by Clarence H. Danhof, Thomas S. Berry, and George Rogers Taylor. For an excellent example of the method of testing such reconstructions, see Richard A. Easterlin's comment on the paper presented by Professor Gallman at the Conference on Research in Income and Wealth held at Chapel Hill in September of 1963. The comment will be published in a forthcoming volume of *Studies in Income and Wealth*.

<sup>8</sup> *J.P.E.*, April, 1958.

<sup>9</sup> *Ibid.*, p. 114; and John R. Meyer and Alfred H. Conrad, "Economic Theory, Statistical Inference and Economic History," *J. of Econ. Hist.*, Dec., 1957, p. 539.

transportation medium.<sup>10</sup> However, the unavailability of the data required to perform such a computation leads Ransom in another direction. He computes the maximum transportation cost that a given commodity could bear after the construction of the canal. The difference between this figure and the actual cost of transporting a unit of the commodity to market is termed the rent per unit of commodity shipped from a particular point, given the existence of the canal. Professor Ransom then computes, in a similar manner, the rent per unit of commodity that would have existed in shipping the specified item from the specified point in the absence of the canal. The difference between these two rents is called the "net rent" attributable to the canal. The net rent summed over all commodities and all shipping points is Professor Ransom's estimate of the unpaid benefit of a canal.<sup>11</sup>

Regardless of whether the central task of a particular study is the reconstruction of missing data, the construction of new measurements or the indirect measurement of a process, theory plays an integral part in the quantitative work of the new economic history. Theory enters, first, in the determination of what it is that needs to be measured. Temin, for example, argues that Hunter's attribution of the western delay in the production of pig iron with mineral fuel to the nature of the region's demand for iron implies that the price of anthracite iron should have been less than the price of charcoal iron. This leads Temin to a search for price data, and the data he finds indicate that the price differential implicit in Hunter's hypothesis did not exist. Temin then argues that the slowness in the western shift from charcoal to mineral iron might be explained not by a change in demand but by a change in the condition underlying supply. Theory enables him to deduce that if the latter process were at work, one would expect to observe that the price of coke iron was initially below that of anthracite iron and then rose toward equality with the price of anthracite iron. The analysis leads to a new search for price data. The data discovered by the search tend to confirm Temin's hypothesis.<sup>12</sup>

Theory enters even more obviously in the attempt to measure indirectly those things which cannot be measured directly. For indirect measurement presupposes a systematic relationship between something that can be measured directly and something which cannot. Ransom uses land values, distances between markets, transportation rates, prices, and quantities shipped—all of which can be directly determined—to measure the unpaid benefits of canals. The link between the available data and the information that Ransom cannot directly observe is the theory of price, particularly the theory of rent.

In this connection, it should be noted that the reliance of the new economic historians on theory is fostered by their desire to determine the extent of the contribution of particular changes in economic institutions, in factor

<sup>10</sup> An example of the application of this approach is contained in Robert William Fogel, *The Union Pacific Railroad: A Case in Premature Enterprise* (Johns Hopkins Press, 1960), Chap. 4.

<sup>11</sup> Ransom, *Government Investment*, Chap. III. Ransom's implicit assumptions are such that his model probably yields a lower limit on the true "net rent." Consequently, his estimate cannot be used to reject the hypothesis that canals were a socially profitable investment. However, Ransom's version of the net rent together with his upward biased "social saving" sets bounds on the true social return to canal investment.

<sup>12</sup> Temin, *A History*, Chap. III.

supplies, in technology, etc., to the observed growth of an economy. The determination of the contribution of such factors to growth involves not only the measurement of events that occurred and conditions that existed but which, because of some "accident," passed unobserved; it also requires the measurement of events and conditions that could not have been observed because they never occurred or never existed. When Ransom asks what the alternative rent on a particular commodity produced in the neighborhood of the Ohio Canal would have been in the absence of that waterway he is posing a question of the counterfactual conditional type. Since the alternative conditions he specifies never existed, the rent associated with these conditions could not have been observed. The determination of the alternative rent must therefore rest on a hypothetico-deductive model.<sup>13</sup>

From the foregoing considerations it seems clear that what is most novel and most important in the new economic history is not the increased emphasis on measurement but the reliance on theory to measure that which was previously deemed unmeasurable. To the extent that the new economic history is successful in statistical reconstruction and indirect measurement via the route of theory, the old confines of data limitations are pushed out and many processes which in the past were deemed unmeasurable become measurable; many issues that once seemed indeterminate become determinate.

It follows that the most critical issue in the work of the new economic historians is the logical and empirical validity of the theories on which their measurements are based. The central importance of the proper specification of the estimating models is revealed in Professor Ransom's criticism of Harvey Segal's estimate of the social profitability of canals. That criticism centers on the period of time used to amortize the capital invested in canals. Professor Segal adopted a fifty-year time span. Ransom suggests that a period of twenty years is more nearly appropriate. The difference between the alternatives may be sufficient to shift the judgment of the social efficacy of the construction of certain of the canals. Which time period should be used depends, as Ransom points out, on the question to be answered. An estimating model based on the fifty-year period is appropriate to the analysis of issues that rest on the social rate of return that would have prevailed in the absence of railroads. However, the judgment of the efficacy of canal construction under the circumstances that actually prevailed during the nineteenth century depends on the actual social rate of return. The computation of the latter rate has to take into account the fact that the advent of railroads resulted in the technological obsolescence of many canals well within the fifty-year period.

The possible consequences of errors in model specification are also illustrated in Professor Fishlow's attempt to estimate interregional trade flows in the antebellum era. Professor Fishlow's computations are based on the assumption that virtually all trade from the West to the South took place through New Orleans. This single-route model of West-South trade affects the computation of the numbers in three of the six cells shown in each of

<sup>13</sup> For a further discussion of the role of hypothetico-deductive models in economic history see Meyer and Conrad, "Economic Theory"; and Fogel, *Railroads*, especially Chaps. I and VI.

the six yearly matrices of interregional trade flows presented in his third table. As a consequence of limiting himself to the New Orleans data in determining the extent of West-South trade, Fishlow concludes that western commodities consumed in the South were a relatively unimportant and a declining share of the shipments from the West during the decades leading up to the Civil War.

Available evidence suggests that Fishlow's findings may be the result of an incorrect specification of the model of antebellum interregional trade routes. It seems likely that during the 1850's a sizable share of western products was shipped to the South via the North Atlantic port cities of New York, Philadelphia, and Baltimore. In the comments that follow I want to focus particular attention on routes of trade between the West and the South Atlantic States of Virginia, North Carolina, South Carolina, and Georgia.

The earliest report readily available to me on the magnitude of the shipments along competing routes from the West to the south Atlantic states refers to the post-Civil War period. According to data collected by the Southern Railway and Steamship Association for the year 1881-82, the overwhelming proportion of southern bound western goods reached the major South Atlantic cities by way of New York and Baltimore rather than by direct rail connection with the West. In that year direct shipments from Chicago to Charleston, South Carolina, came to 4,511 tons; but more than three times that amount was shipped to Charleston through the ports of New York and Baltimore. The ratio on goods sent from Chicago to Savannah was nearly two to one in favor of the northern trunk-line and coastal route.<sup>14</sup> The report of the Southern Railway and Steamship Association leaves little doubt that this indirect route to the south Atlantic states enjoyed a cost advantage that enabled it to dominate—at least in the two decades following the Civil War—all other alternatives:<sup>15</sup>

In making through rates from Chicago to Southern coast points, we take as our basis the trunk lines' rates in effect from Chicago to Baltimore, plus the steamships' rates from Baltimore to Charleston and Savannah; and the totals thus made have been for years, and will probably continue to be, lower than any per mile rate used by the all-rail lines from the West to interior Southern points, as Atlanta, Macon, Augusta, &c. The all-rail lines have therefore to adopt the lowest total produced by combination of trunk-line rates plus steamer rates, adding thereto only a sufficient amount to cover the cost of insurance and allowance for the greater amount of time consumed in transportation by rail and water lines. Southern port rates having thus been arrived at, the rates to Augusta, Macon, Atlanta, &c., have to be made competitive with the Southern port rate, otherwise the business of interior points would be diverted entirely to the coast.

What is most interesting about the Association's statement is the absence of any reference on shipments through New Orleans as a factor in establishing rates from the West to South Atlantic cities. Was this pre-emption of the supremacy of the route on which Fishlow places so much emphasis a phenomenon of the post-Civil War period or did it take place in the antebellum era? Data on relative prices suggest that the usurpation was probably fairly

<sup>14</sup> Charles A. Sindall, "Development of the Traffic Between the Southern States and the Northern and Northwestern States," U. S. Treasury Department, *Report on the Internal Commerce of the United States, 1886* (Washington: GPO, 1886), Tables 4 and 6.

<sup>15</sup> *Ibid.*, p. 683.

complete by 1850, if not sooner. Table 1 shows that while New Orleans enjoyed a price advantage over New York on four basic commodities in the early 1830's, by 1850 prices were lower in New York on two of the four original items as well as on corn. The movement against New Orleans continued through the decade, and by 1859 wholesale prices at New York were from 2 to 39 per cent below New Orleans prices on all five commodities.

TABLE 1  
WHOLESALE PRICES OF SELECTED COMMODITIES AT NEW YORK AND NEW ORLEANS  
(In Dollars Per Unit)

Commodity and Unit	1 New York	2 New Orleans	3 Premium at New York	4 Column 1 as a Percent of Column 2
1834				
Flour (barrel).....	4.98	4.64	0.34	107
Mess beef (barrel).....	9.57	9.29	0.28	103
Mess pork (barrel).....	14.29	12.64	1.65	113
Lard (pound).....	0.079	0.069	0.01	114
1850				
Flour (barrel).....	5.51	5.05	0.46	109
Corn (bushel).....	0.630	0.647	-0.02	97
Mess beef (barrel).....	8.29	11.43	-3.14	73
Mess pork (barrel).....	10.70	10.71	-0.01	100
Lard (pound).....	0.0642	0.0638	0.0004	101
1859				
Flour (barrel).....	5.12	5.54	-0.42	92
Corn (bushel).....	0.863	0.970	-0.11	89
Mess beef (barrel).....	8.84	14.46	-5.62	61
Mess pork (barrel).....	16.49	16.80	-0.31	98
Lard (pound).....	0.108	0.114	-0.006	95

SOURCE: Arthur Harrison Cole, *Wholesale Commodity Prices in the United States 1700-1861, Statistical Supplement* (Harvard Univ. Press, 1938).

Consequently, even if transportation costs between the south Atlantic ports and New York in the late fifties were the same as those between the South Atlantic ports and New Orleans, it would have paid for southern merchants to have purchased in the North. Yet several factors suggest that shipping costs favored the northern alternative. One factor is distance. New York's advantage over New Orleans is 417 miles in the case of Savannah, Georgia; 531 miles in the case of Charleston, S.C.; 717 miles in the case of Wilmington, N.C.; and 1,205 miles in the case of Norfolk, Virginia (see Table 2). Second, even in the 1830's when New Orleans enjoyed a slender price advantage over the northern port cities on mess beef, Charleston, S.C., appears to have made its primary purchases on this item in the North. In every year of the decade the price series at Charleston listed in Cole's compendium is described as "Bos-

TABLE 2  
DISTANCES BETWEEN SELECTED PORTS  
(In Nautical Miles)

	New Orleans	New York	Baltimore	Philadelphia	Boston
Savannah, Georgia.....	1,122	705	625	672	944
Charleston, S. C.....	1,163	632	552	599	871
Wilmington, N.C.....	1,279	562	482	529	801
Norfolk, Virginia.....	1,499	294	172	257	567

SOURCE: U.S. Coast and Geodetic Survey, *Distances Between United States Ports*, Serial No. 444 (Washington: GPO, 1938).

ton mess." Similarly the description of the Charleston pork series in 1836 is "New York mess, March-December," although the average New York price exceeded the New Orleans average by from \$0.50 to \$2.25 per barrel in four of the last ten months of the year.<sup>18</sup> Third, export prices on flour and pork shipped from Atlantic and Gulf ports show a generally rising level of prices as one proceeds down the coast from Baltimore to Key West, after which the prices fall off (see Table 3).

Relative prices thus suggest that if, in 1860, the South Atlantic States incurred deficits in the production of foods that the West produced for export, these states would have made their purchases primarily in the markets of Boston, New York, Philadelphia, and Baltimore. The question then becomes, "How large might the deficits have been?"

Some notion of the magnitudes that might have prevailed can be obtained by utilizing an approach that Professor Fishlow mentioned but did not pursue: estimation of the deficit from production and consumption data. Following Gallman's method of basing estimates of annual meat production

TABLE 3  
AVERAGE EXPORT PRICES AT ATLANTIC AND GULF PORTS FOR THE YEAR ENDING  
JUNE 30, 1857  
(Dollars per Barrel)

	Flour	Pork
Baltimore, Maryland.....	6.72	18.53
Alexandria, Virginia.....	7.00	
Norfolk, Virginia.....	7.87	20.00
Richmond, Virginia.....	8.10	22.76
Wilmington, N.C.....	7.26	
Charleston, S.C.....	7.85	21.30
Savannah, Georgia.....	7.71	23.00
Key West, Florida.....	9.14	
Mobile, Alabama.....	6.72	
New Orleans, Louisiana.....	6.67	19.72

SOURCE: U. S. Treasury Department, *Annual Report on Commerce and Navigation, 1857* (Washington, 1857), Table 7.

<sup>18</sup> Cole, *Wholesale-Commodity Prices*, p. 258.



on the June 1 stock of animals, and adopting Fishlow's assumption of constant average slaughter weights, it is possible to compute national disappearance figures of beef for both 1840 and 1860.<sup>17</sup> Given the constant average slaughter weight of cattle,<sup>18</sup> it appears that for the nation as a whole, consumption declined by 7 percent between the two dates. If one assumes that the South Atlantic States were self-sufficient in 1840 and that per capita consumption in the region declined at the same rate as in the nation as a whole, then the maintenance of regional self-sufficiency required a production of 783,000 head in 1860. However, production appears to have been only 561,000 head, indicating a deficit of 222,000 head or 558,000 barrels.<sup>19</sup> This amounts to about 64 percent of the beef and cattle that Fishlow estimates were shipped from the West to the North.

Applying the same method to pork, it appears that for the nation as a whole, per capita consumption declined by 33 percent over the twenty years between 1840 and 1860. This odd conclusion is the result of the fact that the number of swine in stock grew more slowly than the population and the assumption of a constant average slaughter weight on swine.<sup>20</sup> If one goes on to assume that the South Atlantic States were self-sufficient in 1840 and that per capita consumption in the region declined at the same rate as in the nation as a whole, then maintenance of self-sufficiency would have required a production of 289,000 tons of pork. However, the estimated 1860 production in the region appears to have been 351,000 tons. Consequently, the assumptions of this computation lead to a second startling conclusion: in 1860 the South Atlantic States were a surplus area in pork production.

The unacceptability of these conclusions is revealed by an examination of trends in corn production and consumption. Corn production increased at an extremely rapid rate between 1840 and 1860, rising by 453,000,000 bushels or 120 percent. Since corn was primarily a feed grain—human consumption did not amount to more than 10 percent of the total crop—trends in per capita production figures are relatively meaningless. One can, however, construct an index of "consuming units," establishing a swine as one consuming unit, and counting each other type of animal as some multiple of the "swine unit."<sup>21</sup> On this basis, after making due allowance for exports and seed requirements, corn consumption per consuming unit rises from a national average of 8.7 bushels in 1840 to 13.6 bushels in 1860—an increase of 56 percent.

<sup>17</sup> Conference on Income and Wealth, *Trends*, p. 49.

<sup>18</sup> In the computations that follow, data on animals stocks, production of corn and population are taken from the Census Reports for 1840 and 1860; the ratio of slaughter weight to dressed weight is taken from G. K. Holmes, *Meat Supply and Surplus*, USDA, Bureau of Statistics, *Bulletin No. 55* (Washington, 1907). A barrel of beef or pork is assumed to have weighed 200 pounds. As previously noted, only four states are included under the designation "South Atlantic States."

<sup>19</sup> Following Towne and Rasmussen, the average slaughter weight of cattle is assumed to be 950 pounds. Conference on Income and Wealth, *Trends*, p. 283.

<sup>20</sup> Following Fishlow, the average slaughter weight of hogs used in the computation was 200 pounds.

<sup>21</sup> The weights used in the construction of this index are based on data in USDA, Agricultural Research Service, *Production Research Report*, No. 21 (Washington, 1958); and Fogel, *Railroads*, Chaps. II and III.

Since no plausible increase in human corn requirements could have accounted for more than a very small portion of the indicated rise in per unit consumption, these figures imply a major change in animal husbandry over the twenty-year period. There was obviously a major movement away from grazing (in the case of horses and cattle) and mast feeding (in the case of swine) to corn feeding. Such a turn suggests a rise in the slaughter weight of hogs and cattle.

Did this change from mast feeding and grazing to corn feeding and the consequent rise in animal weights extend to the South? A recent study by Professor Eugene D. Genovese indicates that it did not. Referring to the period of the 1850's, Genovese states that "the quality of southern hogs was far inferior to the better fed, better bred, better housed hogs of the Middle West" and that southern hogs were "allowed to run wild in the woods and feed themselves throughout the winter as best they could, often receiving no grain at all during the year." According to a sample of plantation records drawn by Genovese, the average slaughter weight of 4,000 hogs on plantations in eight states in the year 1860 was 140 pounds, as contrasted to an average 1860 weight of 228 pounds in Chicago, and about 200 pounds in both Cincinnati and St. Louis.<sup>22</sup>

The data on corn support Genovese's finding. In order for corn consumption in the South Atlantic States to have risen at the same rate as the rest of the nation, the four states in the region would have had to have imported 43,400,000 bushels of corn in 1860. But this figure is 130 percent larger than Kohlmeier's estimate of the amount of western corn shipped to the North in 1860. Indeed, it exceeds his estimate of total western corn shipments to all points outside of the region by 90 percent.

The data on corn consumption and Genovese's findings on average weights imply that the apparent decline in per capita pork consumption in the nation between 1840 and 1860 may be an illusion produced by the failure to take account of the rise in the average weights. Since the national corn consumption per consuming unit in 1840 was somewhat below the corresponding figure in the South Atlantic States in 1860,<sup>23</sup> we may take Genovese's estimate of the average weight of hogs in the South in 1860 as indicative of the national average in 1840. This suggests a rise of 43 percent in the national average slaughter weight of hogs between the two dates (from 140 to 200 pounds) a rise that just offsets the decrease in the stock of swine per capita. It also suggests that the per capita consumption of pork remained stable during the period in question.

If the previous computation of the pork import requirements of the South Atlantic States is altered to reflect a constancy in per capita consumption, one finds that pork production in the region fell short of consumption requirements by 787,000 barrels or about 62 percent of the Kohlmeier-Fishlow estimate of western pork shipments to the north Atlantic area in 1860.

Using Fishlow's prices, it thus appears that if one limits oneself only to

<sup>22</sup> "Livestock in the Slave Economy of the South—A Revised View," *Agric. Hist.*, July, 1962, pp. 147-48.

<sup>23</sup> The respective figures are 8.7 and 10.6 bushels.

beef and pork, western meats and animals amounting to about \$21,000,000 which Fishlow assigns to northern consumption may actually have been destined for the South. But such a shift would significantly alter one's appraisal of the South as a market for western goods. Its share in the internal consumption of western exports would rise from 25 percent to 40 percent.

The preceding estimates are too crude to be more than an illustration of the scope of the issue posed by Professor Fishlow's specification of a one-route model in West-South trade. The resolution of the issue requires research on trends in the average weight of animals, on the shifts in the proportion and amounts of the various grains fed to particular categories of animals, and on the human consumption of particular foods in the major regions of the nation prior to the Civil War—research of a magnitude that could not be undertaken in the limited time available for the preparation of this paper.

Professor Fishlow raises a second issue of model specification when he writes:

On the other side, the southern states were far from dependent upon the agricultural largess of the West for their needs. The greater than average per capita production of corn, peas, and beans in the South supports this observation, so too do the larger cattle and swine inventories. The southern social structure, with its large number of landowners with few slaves or none at all, also lends support. It is suggestive of an economic organization with both widespread self-sufficiency and local sale of foodstuffs to nearby plantations.

This position is not only a consequence of Professor Fishlow's single-route model of West-South trade, and an assumption of constancy in the average weights of livestock in the various regions, but also of the high level of aggregations in Fishlow's overall model of interregional trade. The statement that on a per capita basis cattle inventories in the South exceeded that of the rest of the nation is true in 1860 only if one lumps Texas and Arkansas together with the South Atlantic and Eastern Gulf States. Without the two trans-Mississippi states, the rest of the South had a per capita cattle inventory equal to only 94 percent of the national figure.

Texas, which alone had cattle stock exceeding that of all the South Atlantic States, had only one-seventh of the combined population of these states. It is of interest therefore to ask, "What happened to this enormous cattle surplus?" Fragmentary data suggest that large numbers of Texas cattle were sold in Louisiana and the Gulf States to the east of it during the 1850's. Between 2,900 to 6,000 head per year were reportedly sent by boat to New Orleans from Galveston alone during 1850-56. In the first ten months of 1856 an additional 32,412 "beeves from western Texas crossed the Sabine River into Louisiana." The *Galveston News* reported that most of these animals were being driven to New Orleans.<sup>24</sup>

A different structuring of regions and the elimination of the one-route hypothesis thus appears to reveal a steady decline in self-sufficiency in the South Atlantic and Eastern Gulf States. The relative decline which Fishlow observed in receipts at New Orleans appears to be explained partly by the rise of a new western region between 1840 and 1860—the Southwest—which replaced the Northwest as a supplier of the deficits of the Eastern Gulf States,

\* Wayne Gard, *The Chisholm Trail* (Univ. of Oklahoma Press, 1954), pp. 24-25.

and partly by the cost advantage that enabled the northern route to the Southwest to usurp the hegemony of the Mississippi.<sup>25</sup>

Professor Fishlow's paper thus illustrates the two main issues of model specification that confront the new economic historians. The first issue pertains to the empirical validity of the assumptions that underlie their models; it involves the determination of the extent to which errors in specification will lead to a distortion of the reality that historians wish to reconstruct. In the present case, this problem is illustrated by scope of the error that might be introduced by the assumption that there was only one nontrivial commercial route from the West to the South in the antebellum era. The second issue pertains to the power of a model to yield information of the type required to successfully evaluate a given set of circumstances. In the present case, this problem appears in connection with the extremely high level of aggregation on which Professor Fishlow chose to cast his analysis, a level of aggregation that may be too high to permit one to determine either the direction, or the magnitude or the developmental significance of changes in the degree of local self-sufficiency in the South.

My stress on the distinctive characteristics of the new economic history may seem to imply that the work of the new generation described by Professor North is discontinuous with previous work in the field. It does not. The discipline of economic history was no more monolithic in the past than it is at present. Every one of the elements which taken together serve to define the new economic history can be found in one or another of the classics of the past. Despite the general tendency of economic historians prior to World War I to deprecate the usefulness of classical theory, Guy S. Callender invoked the writings of Mill in his pithy analysis of "The Economics of Slavery."<sup>26</sup> And Eli F. Heckscher's thirty-five year old "A Plea for Theory in Economic History" is well known.<sup>27</sup>

One can also find ingenious applications of indirect measurement in economic history well before the emergence of the era of econometrics. As Carter Goodrich recently pointed out, in the absence of data on per capita income, J. H. Clapham used the prices of rags in various nations to rank their level of prosperity.<sup>28</sup> Synthetic statistical constructs are not incidental to but are

<sup>25</sup> Further consideration must also be given to the possibility that large numbers of cattle and swine were trailed into various parts of the South from the Northwest and thus have been omitted from Fishlow's figures (cf. Genovese, "Livestock," *passim*). To the extent that the southern deficit noted above was supplied in this manner, Fishlow's estimates will be off not only with respect to the division of western exports between the North and the South, but also on the overall magnitude of the exports.

<sup>26</sup> *Selections from the Economic History of the United States, 1765-1860* (Ginn and Co., 1909), p. 741.

<sup>27</sup> Reprinted in Frederic C. Lane, ed., *Enterprise and Secular Change* (Richard D. Irwin, Inc., 1953).

<sup>28</sup> "Economic History: One Field or Two?" *J. of Econ. Hist.*, Dec., 1960, p. 535. It is interesting to note that Abbott Payson Usher characterized Clapham's *Economic History of Modern Britain* as "the first attempt to apply quantitative methods systematically to economic history on a large scale." Usher went on to express the view that the quantitative method furnished "an objective standard" for the explanation of economic growth. "Without statistical control," he said, "the timing of basic changes will be misjudged, and the significance of particular technological innovations is likely to be overemphasized." "The Application of the Quantitative Method to Economic History," *J.P.E.*, April, 1932, pp. 186 and 197-98.

the foundation of such landmarks of economic history as Earl J. Hamilton's three-volume study of money, prices, and wages in Spain, Walter B. Smith and Arthur H. Cole's *Fluctuations in American Business*, and Thomas Senior Berry's *Western Prices Before 1860*.

Although the discipline originated largely as a revolt against classical theory and though it has generally shunned the use of modern statistical techniques, economic history has always harbored scholars with a keen interest in the applications of theory and statistics to historical problems. The new economic history is the contemporary continuation of this theoretical-quantitative tradition, fortified by the methods developed and the experiences gained in recent empirical studies of economic growth, by the increasingly powerful tools of mathematical economics and econometrics, by the increasingly varied models of economic theory, and by the improved understanding of the possibilities that exist for the adaptation of models to the analysis of problems and situations other than those that prompted their invention.

# COMPARATIVE COSTS AND ECONOMIC DEVELOPMENT

## COMPARATIVE COSTS AND ECONOMIC DEVELOPMENT: THE EXPERIENCE OF INDIA

By WILFRED MALENBAUM  
*University of Pennsylvania*

### I. Introduction

Under given market conditions, contrasts are readily made between allocation decisions which emphasize principles of comparative advantage on the one hand and so-called "principles of growth economics" on the other. However, both the theoretical and empirical dilemmas posed by such contrasts can be overstated. On the theoretical side,<sup>1</sup> they seem to stem from differences in assumptions and scope of the two doctrines as typically formulated. Dynamic assumptions can bring harmony and perhaps identity out of the contrasts. Similarly the historical evidence does not make the case for one set of principles as against the other.<sup>2</sup> Rather, the record provides illustrations of both growth and non-growth, each with specialization of output and relative expansion of external trade, or with diversification and relative decline. We emerge only with the important proposition that the appropriate allocation policy for economic development must rest on assumptions which reflect actual possibilities confronting a poor nation, in terms both of its domestic capabilities and of those of the rest of the world.

In other words we best move rapidly from doctrines and principles to descriptions and particulars. What light do the two allocation theories throw upon the growth problems of the Indian economy? India is making a deliberate effort to embark upon a state of growth from a state of stagnation. Allocation decisions are made explicitly to that end. There is much information about plans, somewhat less about their effects. And, finally, Indian history also provides some insights for trade and growth in the preplanning years.

Before the nineteenth century, India already was an important exporter of plantation and other crops. Increasingly, Indian raw cotton found ready processors in England and other manufacturing countries. In the nineteenth century, clearly by the 1850's, rapid growth and diversification took place in India's own capital stock; the products of

<sup>1</sup> See, for example, H. B. Chenery, "Comparative Advantage and Development Policy," *A.E.R.*, Mar., 1961, pp. 18-51.

<sup>2</sup> See C. P. Kindleberger, *Foreign Trade and the National Economy* (1962), especially pp. 177-211.

India's modern industry, notably jute and cotton textiles, entered the world market. The relative importance of manufactures in total exports expanded: in the 1920's they may have contributed 25 percent of the value of all exports. By the beginning of the plan program, this ratio stood at almost 50 percent, mostly as processed cotton and jute. Indeed, India had even become a "modern" raw cotton importer. But neither the gains from trade nor the gains from initiating industrial diversification spread their influences broadly through India's economic fabric. Many reasons can be suggested; none can yet be adequately documented. In some degree, the nature of the specific products, the conflict between domestic and foreign textile producers, the nationality of Indian government—all play some role.<sup>3</sup> In any event, the new government of independent India could look back upon a long trade history in which the value of India's current exports, composed of diverse goods and services, may have exceeded that of imports, despite a stagnant if not deteriorating standard of living in the nation. The direct and indirect gains from trade specialization or from industrial diversification had been neither large nor clear. Imports and exports averaged less than 5 percent of the total economic product in 1948-50. India, along with the other "big" nations like Mainland China, the U.S.S.R. and the U.S., consumed and utilized essentially what it produced. In India (as in China) this situation coexisted with what was probably the lowest average consumption levels prevailing over any significant area of the world.

## II. *India's Planning for Foreign Trade*

The long-term perspective (1951-76) for Indian planning projects a developing economy that is a more self-sufficient economy. This is not said explicitly. It is a statement more documentable from the statistics of the plans than from their rationale. Thus, specific figures in the Second Plan indicate exports of goods and services at a lower ratio of national product in 1960-61 than in 1955-56. A similar trend for exports pertains for the current program.<sup>4</sup> Finally, tentative government projections for the end of the fifth plan (1975-76), when India might no longer need extraordinary foreign assistance, show export trade at a still lower ratio of total output than in 1960-61 (to say naught of 1950-51). As compared with the preplan years, a doubling of exports is foreseen by 1975-76 as against an almost fourfold increase in national product.<sup>5</sup> The expanded export trade is to reflect the new and broader

<sup>3</sup> See, for example, my *Prospects for Indian Development* (1962), Chap. VII.

<sup>4</sup> W. B. Reddaway, *The Development of the Indian Economy*, p. 162.

<sup>5</sup> These rough comparisons are in current rupees, not adjusted for price level. Both trade and output figures for 1975-76 are exaggerated, perhaps not in equal measure. But the argument is not affected.

industrial base of the economy. Agricultural raw materials and their related manufactures (especially tea, jute, and cotton textiles) will constitute about 60 percent of export value in 1975-76, as compared with almost 85 percent in 1960-61. Manufactured metals, engineering products and machinery are to multiply their export value ten times and more (machinery itself, thirty times).

On the side of the rationale of planning, one is encouraged to infer that the chosen path for growth is an optimum, least cost path. Given the specific growth objectives and the social goals of the nation (and some idea of the extent to which grants and loans can be obtained from abroad), the planning task is posed as essentially one of resource allocation and use so that the income flows over time, appropriately discounted, are maximized. Theoretically, the time flows and composition of foreign trade are results, not a priori assumptions, of the planning exercise. Admittedly, India's coal, iron, and other mineral resources, and its long history of technological and managerial accomplishments in modern industry, do provide parameters which result in a major focus on industrial diversification. There is also a factual judgment on demand elasticities which yield limited scope for export expansion. With these "facts," the heavy import substitution of the plans becomes the only course of action—not a selected alternative.<sup>6</sup> If the new industrial base is to permit a level of foreign trade which grows relative to national income, this prospect is for a longer-term perspective—in the uncharted period beyond 1976.

It is of interest to note that actual developments in India did follow what might be termed classic lines. Before the plans, India was importing goods which were on the whole more capital intensive than was India's total product; India's exports on the other hand were relatively more labor intensive than were these imports.<sup>7</sup> With domestic terms of trade more favorable to the high-capital input goods and the international terms confronting India relatively favorable to high-labor

<sup>6</sup> Actually India's plans proceed on the assumption that for political and broad international considerations, India should be relatively independent of foreign sources for goods and services considered essential to the nation's future. At least from the explicit presentation of the so-called "plan-frame" for the Second Plan—late in 1954—a large measure of self-sufficiency in capital goods became an integral part of India's official planning objectives. Thus the Third Plan calls for the establishment of "machine-building capacity so that the requirements of further industrialization can be met within a period of ten years or so mainly from the country's own resources." This orientation has implications beyond those of the time flows and specific composition of outputs and of inputs. It bears also on the savings performance of the society and upon the ability to attract various kinds of foreign investment (perhaps even assistance). Such a projection need no longer trace a least-cost path. Since a full comparison is in actual practice out of the question, it is relevant that the official plans do not even pose these alternatives. They proceed as though the chosen path was the optimum economic path.

<sup>7</sup> For a statistical argument to this effect, see R. Bharadwaj, *Structural Basis of India's Foreign Trade* (1962), pp. 55-59.



input commodities, there were important gains from trade.<sup>8</sup> In the plan years, India's productive potential for capital intensive output has grown relative to that in the rest of the economy. The supply of capital expanded more rapidly than the supply of labor. Rough estimates suggest 50 percent for the former in contrast to an 18 percent growth in the labor force over the past twelve years.<sup>9</sup> And second, it can be inferred that, on the average, more capital-using techniques of production were applied over the period.<sup>10</sup> Incremental capital-labor ratios were planned at progressively higher levels, with larger increases in the capital intensive sectors than in the rest of the economy. All other things equal, this would tend to decrease the ratio of labor that would be employed (at given wage levels) in the more capital intensive sector as compared to the more labor intensive.<sup>11</sup> On both counts—changes in factor supplies and in the techniques of their use—the production frontier in India, it can be inferred, shifted in the direction of the output capabilities of India's trading partners.

India can therefore more nearly achieve the product use pattern of the earlier period without moving as far from its own production frontier. The absolute level of foreign trade can become smaller relative to national product; domestic production thus substitutes for imports. Domestic output of the more capital intensive products grows relative to domestic production of other goods. Insofar as this occurs, domestic and foreign price relationships remaining unchanged, the gains from the policy have been import substitution more than real income. As relative endowments change, if domestic terms of trade also move toward the same ratio as the international terms of trade between the two categories of goods, all "gains from trade" are diminished; there is a loss of real income to be offset by the expanded product from overall, more self-sufficient growth.<sup>12</sup> In this "classic" argument

\* This statement (as also part of the ensuing discussion) is phrased with the familiar diagrammatic representation in mind of a national production frontier, where output is categorized as more labor intensive (for India more export-oriented) goods or more capital intensive (more import-oriented) goods. See, for example, G. M. Meier, *International Trade and Development* (1962), pp. 9-39.

<sup>9</sup> These are rough calculations. New capital formation in India after 1950-51 may have totaled \$28.5 billion in 1948-49 prices, as against a total supply of fixed capital of some \$65 billion at the beginning of the plans. For present purposes it may suffice to estimate that the labor force has increased by 25 million workers above the 140 million or so early in 1951.

<sup>10</sup> Bharadwaj's calculations suggest that this did occur in 1958-59 as compared with 1950-51. (*Op. cit.*, pp. 87-112.)

<sup>11</sup> This relative saving of labor in the first sector did not seem to spur any significant increase in output in the others. The use of more advanced production techniques over the three plans tended to be accompanied by a relative growth in the level of unemployment of various kinds.

<sup>12</sup> In point of fact, Indian prices seem to have moved in the opposite direction, with capital intensive (import-oriented) prices increasing relative to the other prices. This reflects the high relative costs of import substitutions and indeed some of the limited linkages in the Indian economy. (See IV below.)

the case for India's policy of growth must rest on the demonstration that subsequent expansion will be easier in a more self-sufficient nation.

Given India's physical and human endowments, import substitution might well go a long way toward replacing imports as the economy grows. Sooner or later the cost structure of the more self-sufficient nation needs to be tested in a world market, at least in some measure. Given realistic assumptions about this market, there is a point where growth suffers as import substitution expands. Granting India's potential for diversification, how can it be implemented so that long-term growth is aided, not impeded?

### *III. Implementation of Trade Policy*

To diversify its industrial base a developing nation must use more capital goods than even a rapidly expanding domestic output of these goods will permit. In the early years large amounts of capital goods have to be imported along with an increased amount of intermediate products required for operating the new capital. In addition, pressure for more imports of consumer goods can be expected. This requires an initial dependence upon foreign sources, given the relative inelasticity of domestic output in a long-static society. With respect to exports, domestic consumption competes more aggressively with foreign demand, which is generally believed to be relatively inelastic for the "staple" types of exports from poor countries. Hence we find vigor in the import expansion, relative stagnation in exports, and large international deficits on current account.

Even after allowing for the large foreign exchange contribution of grants, loans, and private investment, selection must still be made among these import pressures. Government extends export subsidies of various kinds, licenses new installations, applies a broad complex of import restrictions, and takes selective action on prices. To date, these tools have not sufficed to achieve the "balance" (i.e., the degree of imbalance) planned. The present (unofficial) premium on foreign exchange probably exceeds the official rate by 40 to 50 percent; by and large domestic production costs for many import substitutes are at least that much above the landed cost of equivalent foreign goods. The level and the pattern of actual imports thus differ from what the planning authorities envisaged. This in turn influences the planned program of import substitution.

Planning such a program poses difficult technical problems, even for India whose statistical resources are so much better than those in other poor nations. Import substitution goes beyond replacing direct imports of the substitute goods. Domestic output at higher cost means greater money incomes, perhaps greater pressures for imports of consumer

goods. Home production in one industry will step up requirements in other domestic industries which depend on imports to some degree. To my knowledge there is not yet in hand an analytic (input-output) table which can approximate total imports associated with a stipulated major expansion of output to substitute for purchases from abroad. Similarly, data are not adequate to predict the import demand under the types of income changes now occurring in India. In addition to these technical considerations, there are the complex problems of administration of such controls, particularly in a country where a relatively free economic system has long prevailed. Delays in obtaining certain imports may raise production costs; they may also provide goods no longer needed by the original purchaser and which thus become available for resale for purposes other than those originally authorized. Data to illustrate these problems of implementation are not available, although the Second Plan does yield some insights. Thus imports during 1956-61 totaled more than \$11 billion, some 20 percent above the plan projections. Of this excess (more than \$2 billion) some 10 percent was due to an underestimate of the actual imports of machinery and equipment needed for capital formation. Imports of consumer goods (including food grains financed under P.L.480) exceeded plan by about \$750 million. The largest excess (almost \$1.2 billion of the underestimate of some \$2 billion) arose in the category of raw materials and intermediate goods required to maintain output in the expanding industrial sector.<sup>13</sup>

The experience of these years has prompted a greater attention to problems of maintenance imports for the Third Plan. Even so, the Planning Commission in 1961 expressed doubt that its increased provisions for maintenance imports during 1961-66 would be sufficient: "some underutilization of capacity will have to be tolerated" in the Third Plan (*Third Plan*, p. 110). With half of the plan now a matter of record, total imports are more or less as scheduled, particularly so the "direct" substitute imports of plant and machinery. But the rate of national growth through 1962-63 was less than half of what the plan projected; the increase in industrial output proper was some 60 percent of the target. Among the reasons for this, import restrictions for goods "indirectly" associated with import substitution loom large. It is clear that solutions for the technical and the administrative problems of "what imports" for import substitution are not yet fully in hand. In addition, India's import policy has as yet not reduced the ratio of total imports to national product.

<sup>13</sup> The import record must also be interpreted against the fact that there was a reduction in 1958 of the plan's projections for total output, and in particular for the expansion of industrial capacity. Even so, output in the last years of the Second Plan may have suffered from a shortage of the critical maintenance imports.

Import substitution, especially where new governmental industrial undertakings are of key importance, seems to have some proclivity for high costs, which in turn delay the date when competitiveness in the world market can be achieved. Foreign aid to India has provided a large part of the resources for these undertakings; some 75 percent of this aid takes the form of project assistance as against "free exchange." Tied loans and grants tend to be expensive loans and grants. This problem is posed in what may be debatable terms of national effort: are the vast grain shipments under P.L.480 and India's lagging grain production related in any way? Not debatable is the truth that shipments under foreign assistance programs cost appreciably more than what free exchange could purchase on the competitive international market. Arguments that this is of small significance, given the high premium India places on foreign exchange, the high rate of discount of the future, and indeed the fact that the imports may be "free," fail to provide any answer for the enterprise to which the equipment goes. Capital and maintenance costs (for replacement and parts) tend consistently to be above those in comparable firms, perhaps in other countries, using similar if not identical equipment purchased in the free market.

In some measure, high costs for substitutes escalate to other parts of the economy. Thus rubber tires made by a foreign company which always imported rayon fibres have now substituted domestic rayon fibers. For comparable qualities, rayon costs have increased by some 35-50 percent. Or consider the new synthetic rubber plant to which this company increasingly turns for its basic raw material. The new plant was constructed to use indigenous alcohol, the supply of which depends on the output of sugar cane and the flow of sugar to refineries as against the domestic gur producers. To date, supplies of domestic alcohol have limited rubber output to well below plant capacity. (Future alcohol supplies will continue to depend upon governmental policy on prices for different cane users.) On both accounts synthetic rubber prices are pushed higher relative to prices for imported rubber.

To some extent, the import substitution emphasis encounters higher costs which stem from other aspects of government policy. Thus many of India's new industrial plants in the private sector are smaller than they should be if they are ever to have efficient low-cost operation (petrochemicals, for example). The reasons for limits on scale may be traceable to a socialistic policy concern about very large (monopolistic?) enterprises. Similarly, new plant locations are influenced by political considerations, with consequent transport and other diseconomies (a point well illustrated again by our synthetic rubber factory). With respect to the public sector enterprises directly, such information

as is available indicates that their costs have been higher than those of comparable companies in the private sector.<sup>14</sup> However temporary such results, they contribute to India's present difficulties in making import substitution economic. In the same broad category are recent efforts to improve the efficiency of the railways (a powerful national enterprise). "Block rakes" for coal delivery have imposed additional cost burdens on collieries on the one hand and on coal users on the other. Again, this situation will eventually work itself out, but in the meantime the railway's gain may well be offset by higher costs elsewhere in the economy.

None of the above points constitutes an argument against a major focus on import substitution. Progress will be slow, and eventual competitiveness in many lines will be long delayed. Few of India's infant industries have outgrown tariff protection throughout India's long modern industrial history, and the circumstances under which new enterprises are now being established give little promise of change in this respect. Perhaps a greater emphasis on relative costs and efficiency, as well as on substitution, may facilitate the transition.<sup>15</sup> Somehow or other realistic shadow rates of exchange and of interest must begin to influence cost estimates in the private sector and in the public sector. (As of this time, such rates would price the rupee at fourteen cents at the highest; capital would return 15 percent at the lowest. The outlook in future years may warrant considering these as minimum adjustments.) The need for efficient import substitution must be given economic content. India is a nation where private entrepreneurs have had long experience as traders and financiers; the present high money income situation in the industrial sector begets new forces which may adversely affect the prime objective of economic growth. Thus the scarce licensing and foreign exchange allocations have become a profitable source of earnings in themselves. The recent developments in the steel industry, for example, are giving rise to remedial action which, however meaningful in eliminating the profitability of dealing in licenses and exchange permits, may impair the very allocation structure for which steel expansion in the public sector was sponsored.

A new impetus to market forces can be expected also to strengthen the growing realization that export stimulation may have been neglected in the emphasis on import substitution. Thus, recent analysis, necessarily tentative in the nature of the subject, lends credence to the con-

<sup>14</sup> See Forum of Free Enterprise, Bombay, "A Survey of State Enterprises in India," May, 1962.

<sup>15</sup> In a recent comment on government's import policy, the influential Federation of Indian Chambers of Commerce and Industry argued that in two major categories (largely intermediate goods and components, spares, etc.) "... we can very readily programme to manufacture at least goods worth Rs. 500 cr., which is nearly 50% of our total [1960-61] imports." Only those goods should be imported that cannot be duplicated; relative costs are not considered. (F.I.C.C.I., "Imports and Industrial Development," 5 June 1962.)

clusion that India may have allowed itself to be priced out of key markets for important exports.<sup>16</sup> In contrast to the frequent claim that India confronts a relatively inelastic world demand for key export commodities, there is evidence that India has been losing its proportionate share of the world market. This is true especially for goods for which world demand has grown—the situation which actually characterizes the position for most of India's exports (70 percent or so) in recent years. If India had maintained its share of these markets, export proceeds in recent years would have been larger, by 10 percent or more.<sup>17</sup>

#### IV. *Foreign Trade and Economic Growth*

We have been concerned with foreign trade policy and its implementation. Foreign exchange (and other) problems limit India's import potential: hence domestic sources for these goods, initially at higher costs that will eventually come down. But it would be a mistake to assume that when (if) India provides itself with capital goods in sufficient quantity, perhaps as stipulated in its plans, this will solve both the trade and the development problem. As I indicated above, the trade problem (for India, the trade-substitution problem) is only at the beginning of being solved. Even when solved, however, India's major development tasks might well remain. The truth is that all of India's development effort to date, and all those now projected through the fifth plan (1975-76), provide little assurance that India's needs in agriculture or in employment will be met. The most successful achievement in industrial expansion (including the trade success above) can make but marginal contributions in these critical areas. The present focus of planning authorities on the modern industrial parts of the economy must continue, but must be matched by additional major efforts in agriculture and in employment. Indeed, one might say that without progress in agriculture and employment generally, the best industrial (and trade substitution) efforts will be thwarted. This certainly is the lesson from India's past experience—viewed, as planning officials in India prefer, at this midpoint in the long period perspective of twenty-five years. With all credit to the past achievements, under a democratic form of government, there are few who would objectively regard the paths to future growth as well established. On the contrary, there is a growing appreciation, in operating as well as in scholarly circles, that new directions are very much needed, and sooner or later to be expected.

The new effort must be characterized by particular attention to the

<sup>16</sup> For tea, cotton cloth, manganese ore, cattle hides, and goat skins India's prices relative to competitors seem to be associated with India's declining share of the U.K. market. See a forthcoming article on Indian exports by Benjamin Cohen.

<sup>17</sup> Cohen, *ibid.*

combination of abundant manpower and limited nonhuman resources; it must supplement and not substitute for the present effort, with its emphasis on industry and related services. Programs are needed to assure an upward trend in agricultural output—at least of 4 percent per annum over the next two or three plan periods. Programs are also needed to assure the continued growth of overhead capital which can be constructed with relatively heavy labor inputs. There is wide scope for output of additional labor intensive consumer goods for the domestic market, as well as for international trade. In a true sense, India's output in the future must bear more evidence of India's factor endowments. In both remote and recent past, neither India's traditional foreign trade leading sector nor its modern industrial leading sector has provided a path for growth. In the real solution of the growth problem that still lies ahead for India, proponents of growth economics—"push diversification, push industrialization"—will be able to support their views. But the condition for this is a greater reliance by India upon its relative resource endowments which will, for decades to come, continue to delimit India's comparative cost structure.

## COMPARATIVE COSTS AND ECONOMIC DEVELOPMENT: THE AFRICAN CASE

By WALTER A. CHUDSON

*United Nations*

### I. *The Historical Record*

If foreign trade served in the nineteenth century as an engine of growth for the majority of countries in Latin America, Asia, and the "regions of recent settlement," the history of modern African economic growth—which, with the exception of the Republic of South Africa and a few other countries can be dated from about 1900—is a tale of exports *par excellence*. The expansion of exports from Africa has been the mainspring not only of its economic growth but of the commercialization of its economic activity in general. It has thus promoted economic development in its two aspects: growth and change.

If this assertion be accepted, there still remains the ready response: "Yes, the engine of such growth as there has been." And the possible retort: "Yes, and of such growth as might reasonably have been expected."

A glance at the economic map of Africa illuminates the first response. At the end of the 1950's, three countries—Egypt, Algeria, and the Republic of South Africa, the modern economic history of which substantially antedates that of tropical Africa—accounted for about half the value of the region's exports, including gold, and 40 percent if gold exports are excluded. These three countries represent about 17 percent of the region's population. If attention is limited to sub-Saharan Africa, the Republic of South Africa accounts for about one-third of the total value of exports, including gold; adding the Congo (Leopoldville) and the Federation of Rhodesia and Nyasaland raises the share of these three countries to about 60 percent of the total with about 22 percent of the population.

Since national incomes in African countries are highly correlated with exports, it is tempting to conclude that even under more favorable world market conditions for many primary commodities than may be expected to prevail in the next decade or so, the export expansion during the period 1900 to 1960 in Africa was a lop-sided and spotty affair geographically and that the presumed engine of growth cut a rather haphazard swath through the continent; to which it might be added that the burden of proof rests on those who would assert that a better dispersion of growth through export expansion may be ex-



pected in the future. But the balance sheet cannot be quite so simply drawn.

It is true that the center of gravity of the export expansion (leaving aside Egypt and Algeria) was in a few mineral producing areas, typical of the "enclave" economy. In most other areas, however, the export expansion, while less important in regional terms, was largely based on agriculture and more diffused in its impact. This was true particularly where the expansion was based on small-scale production (as, for example, in Ghana, Nigeria, and Uganda), but even some plantation-type operations, such as Sudanese cotton and Tanganyika sisal, had fairly far-reaching repercussions on hitherto static economies, particularly when account is taken of their impact on the labor market, which extended frequently beyond national borders. The development of agricultural export commodities was the primary force transforming subsistence production into production for the market. And it is important to add into the balance the contribution of this transformation to the commercialization of agriculture and of ancillary services for domestic consumption. For this reason export expansion through agriculture may have had a stronger catalytic effect on productivity than mineral-based expansion, particularly in areas where the peasant producer was predominant. But the transforming impact of mineral-based expansion on the labor market should not be neglected.

When we think in terms of tropical Africa (and that is the area with which the rest of this paper is largely concerned), it becomes apparent that, without denying the dominant role of exports in development thus far, their growth has been far from spontaneous. It is true that the coffee plant was found growing wild in Ethiopia, but many of the crops on which the export economy of tropical Africa is largely based, as well as some important food crops, are not indigenous. Furthermore, the development of export crops has been fostered at considerable cost by "infant industry" policies.

This is far from saying that an equal amount of subsidization of industry would have produced an equivalent or superior result at that stage of African development. It seems doubtful whether a strong case can be made that Africa has suffered appreciably from "overexporting" in the sense defined by Arthur Lewis, particularly when account is taken of the small share of world supply which Africa's exports have represented until recently and of the fact that the expansion was, with some notable exceptions, based on voluntary small-scale peasant production.<sup>1</sup> But the cost of the export expansion and its impact on the structure of production must be acknowledged. In the case of the French-speaking territories (and of former Italian Somaliland), the cost has included

<sup>1</sup> W. A. Lewis, *The Theory of Economic Growth* (London, 1955), p. 281.

not only the usual overheads of agricultural extension, transport development, port and storage facilities and the like but also a substantial subsidy in the form of sheltered export prices in the metropolitan market, even after allowing for the effect of a sheltered market in the territories for imports from the metropole. And it is these countries which remain the most dependent on external aid at their current levels of expenditure on administration, education, health, and the like.

A more spontaneous growth of exports occurred in certain other areas; for example, Ghana, Nigeria, Kenya, Uganda, and the Rhodesias. In general, these are areas that "paid their way" to a considerable extent during the 1950's and where an industrial expansion largely based on import-substitution is more advanced.

What has been reviewed thus far may perhaps be said to correspond to the opening phase of export expansion in other continents identified by Myint.<sup>2</sup> It has conformed, in general, with a Heckscher-Ohlin model of an economy with a comparative advantage in commodities absorbing a relatively large input of natural resources and, in some cases, labor, the development of which has been triggered by an infusion of external capital, administrative and managerial skills and technology. The question remains to what extent the opening phase may be said to have ended. A confrontation of export expansion and import substitution possibilities—however tentative—is therefore called for.

## II. *Export Expansion*

A useful point of departure is the expansion of African exports since the end of the second World War. In current prices, African exports (excluding gold) rose from \$3.7 billion in 1950 to \$6.6 billion in 1962.<sup>3</sup> The exports of tropical Africa, which now account for about 60 percent of the total, underwent a similar expansion, which was close to the growth of world trade as a whole. This is an impressive performance, particularly in comparison with the sluggish growth of exports from Asia (excluding Japan) and with the virtual stagnation of Latin Amer-

<sup>2</sup> H. Myint, "The 'Classical Theory' of International Trade and the Underdeveloped Countries," *Econ. J.*, June, 1958; and "The Gains from International Trade and the Backward Countries," *Rev. of Econ. Studies*, XXII, 2, pp. 1954-55.

<sup>3</sup> This section draws on two papers by the author in course of publication: "Trends in African Exports and Capital Flows," *Proceedings of the Conference on Indigenous and Induced Elements in the Economies of Sub-Saharan Africa*, Northwestern Univ., 1961; and "The Outlook for African Exports," *Proceedings of the Nyasaland Economic Symposium*, July, 1962. The statistical sources include, P. Lamartine Yates, *Forty Years of Foreign Trade* (London, 1959); United Nations, *Monthly Bul. of Statis.*, July, 1963; United Nations, *World Economic Survey*, 1962, Part I; United Nations Economic Commission for Africa, *Background Paper on the Establishment of an African Common Market*, document E/CN.14/STC/20, Oct., 1963, and *Intra-African Trade*, document E/CN.14/STC/20/Addendum 1; *OECD, Statistical Foreign Trade Bulletin*, June, 1963. Unless otherwise indicated, the data refer to continental Africa.

ica's exports. Furthermore, the recent average rate of increase in the quantum of Africa's exports (5.5 percent per annum in the period 1953-62) has been over twice the rate during the preceding four decades (2.5 percent during the period 1913-53).

This rosy picture must, however, be qualified in two respects. First, the capacity of these exports to finance imports has been reduced somewhat below the potential implied in the figures by a deterioration in the terms of trade of most countries (though measurement of the "loss" is difficult in view of the changed composition of imports during the period). Second, while the value of Africa's exports increased in practically every year of the period, the greater part of the expansion took place before 1957—a period when export earnings were augmented by rising prices. Subsequently, the rate of expansion of export proceeds has been slackening (prior to the recovery of the prices of a number of commodities during 1963) because an increase in the quantum of exports has been substantially offset by declining prices. Almost all countries shared in both the expansion and the slowdown.

Much valuable work on global projections of demand for and supply of agricultural commodities has been done in recent years, particularly by the FAO.<sup>4</sup> Such projections do not, of course, indicate what Africa's share of the totals may be. However, by considering the projected demand for Africa's staple export commodities, some idea can be gained of the minimum prospects on the not unreasonable assumption—on grounds of both comparative advantage and probable international economic policies—that Africa's share should, at the least, not decline.

What emerges is a projected growth, from 1957-59 to 1970, of import demand of "high income" countries (including Eastern Europe) for agricultural commodities exported by "low income" countries in the range of between 2 and somewhat more than 3 percent per year, depending on the optimism of the assumptions. For certain commodities, which bulk large in African trade, the projected expansion is higher; for example, coffee (3 to 4 percent) and cocoa (4 to 5 percent). For other groups it is lower, notably fibers and oilseeds, where a slight contraction is foreseen. Inclusion of projected Eastern European markets in the estimates affects mainly the outlook for cocoa; without that outlet, the prospect is significantly less favorable. The estimates do not take into account intra-African demand, to be considered below.

It is instructive to note the difference between the position of cocoa and coffee, on the one hand, and oilseeds and fibers on the other. Cocoa and coffee are favored by a higher income-elasticity of demand. On the

<sup>4</sup> Food and Agriculture Organization of the United Nations, *Agricultural Commodities—Projections for 1970*, Special Supplement to the *FAO Commodity Rev.*, 1962 (Rome, 1962).

supply side, they are less vulnerable to synthetic substitutes and to the conventional assumption of comparative cost theory of production with relatively fixed factor proportions.

Adding a price dimension to the projections, let alone allowing for the terms of trade, increases their hazardousness. Recent international action with respect to coffee and potential action in other directions suggest that it may be realistic to assume that world prices of Africa's agricultural exports may not decline appreciably below prevailing levels during the period in question (leaving the prospect of a further mild deterioration in the terms of trade). An expansion of meat exports (and in one or two cases, the tourist trade) could substantially alter the prospect for several countries, in both external and intra-African trade, but this important potential development must be viewed in terms of a longer run than one decade. Thus, the high rate of growth during the 1950's of agricultural exports and in their import capacity will probably not, in general, be repeated during the next decade or so.

For several countries fortunate enough to be significant exporters of minerals or probably on the verge of becoming so, the position appears more promising, particularly when account is taken of potential increases in the degree of processing as domestic energy resources are developed. Some contribution can also reasonably be expected from a higher degree of processing of agricultural commodities, a development which has thus far been hampered both by commercial policies, particularly in Europe, which frequently discriminate progressively against more processed imports, and by technological and other limitations in the exporting countries. For other types of manufactured goods, a significant expansion of extra-African trade will probably have to wait upon the growth of production of such commodities for the domestic market, to be considered below.

The trends and prospects sketched above are already reflected in the official development plans of a number of African countries; on the average, they assume a smaller rate of increase in export earnings than in the recent period.

A feature of Africa's export development that deserves explicit mention at this point is Africa's share in world exports of particular commodities and its implications for price movements. The high rate of export expansion described above was based on levels of output and exports which were, except for a few commodities, a minor or even negligible share of the world total. Continental Africa's share in world exports is still not quite 6 percent and the share of tropical Africa about 3 percent. By now, however, Africa's share in world exports of a number of commodities is appreciable and the list is growing, particularly if account is taken of different grades of commodities such as cotton and

coffee (and of the fact that the statistics of world trade include substantial re-exports of commodities originating in Africa).

In the early stages, the effect of the expansion of their own exports on prices could be largely ignored by African producers and governments, and the threat of what has been unbeautifully called "immiserizing growth" through African export expansion could be ignored. This is no longer necessarily true; further, the situation has been rendered more delicate by the position of producers in other less developed countries outside Africa. Part of the ease of African expansion in the 1950's can be attributed to the lagging recovery of production in competitive areas like Indonesia (sisal, oilseeds).

The impact on prices has been manifested most dramatically in the case of coffee. The international coffee agreement, in which the African exporting countries now participate, has met the threat of possible "immiseration" but at the expense of export quotas. The balance of advantage or disadvantage in participating in the coffee agreement has been a particularly painful one for African countries to strike just because they probably are, in terms of comparative advantage, low-cost producers. But the limited capacity of non-African producers to transform their economies in the short run by reallocating resources from coffee production has placed serious limits on Africa's capacity to exploit this advantage. In this sense, development, like peace, is indivisible. Even if absolute "immiseration" is not a threat, a country may fail to get the maximum benefit from its position of comparative advantage because of the sluggish transformation of competitors which have a different set of comparative advantages, at least potentially. If the commodity is subject to an international agreement, the degree to which a country fails to get the maximum benefit may depend, among other things, on its bargaining position at the conference table. Appraisal of the true long-run position is clouded, furthermore, in the case of tree crops—which are a major source of Africa's agricultural exports—by cobweb-type supply reactions of medium-term duration. In pursuit of an optimum allocation of resources a country must still consider the advisability of a "second-best" solution. The comparative productivity of resources for the domestic market remains to be explored.

### III. *The Domestic Market, Import-Substitution and Intra-African Trade*

Starting from its still slim base, production for the domestic market is bound to play a growing role in the economy of tropical Africa. This is true a fortiori if we include within the concept of the domestic market the latent possibilities of intra-African trade. The problem here is to

single out such meaningful comments as can be made without entering fully into a discussion of investment criteria in comprehensive development planning.

To begin with, it would be misleading to think of import-substitution and production for the domestic market generally in terms of industry alone. Something like 17 percent of African imports consists of food, beverages and tobacco products, and the figure for tropical Africa alone may well be higher. With due allowance for the fact that this partly reflects the dual economic structure of many countries with its effect on imports through the consumption patterns of expatriate personnel, it does not seem likely that growing "Africanization" will change the pattern of import demand radically. A good deal of the food imports, furthermore, are relatively unprocessed. On the supply side, it is a well-known and important fact of the African economy, in contrast particularly with Asia, that the scope for a fairly rapid increase in agricultural productivity is great in the sense that there is far less population pressure, with certain exceptions, and that relatively simple changes in growing and processing techniques could yield big increases in productivity. Particularly if account is taken of the potential scope for intra-African trade, the comparative advantage of tropical Africa in domestic food production must, in general, be regarded as strong.

African industrial production (apart from activities directly linked with production for export) thus far has typically followed the pattern of import-substitution designed to satisfy an internal market based on an expanding export sector and augmented in varying measure by rising domestic food productivity. In addition to the growth of *in situ* services and construction, production for the domestic market has consisted largely of an increasing list of consumers' goods (beer and other bottled beverages, processed foods, cigarettes, shoes, cotton textiles, simple furniture, and other household goods). In the category of intermediate and capital goods, the most significant development, apart from energy production, has been the growth of building materials, notably cement, timber, and clay products and some metal fabrication of containers, building components and the like.

Most of this growth has occurred since 1945; in a few instances, an earlier stimulus came from protection due to wartime shipping and supply shortages. A closer examination, of course, reveals significant intraregional differences in the extent of industrial development. Furthermore, an input-output table (of the type constructed by Peacock and Dosser in their study of the national income of Tanganyika) will reveal that the import content of a number of these manufactured goods remains quite large, particularly if account is taken of imported managerial and technical personnel. But a start has been made.

A large number of the products mentioned above are "market-oriented" goods having a high degree of natural protection against imports because of high shipping costs, perishability and similar reasons.

The dependence of African governmental revenues on comprehensive import duties (presently in the range generally of 15 to 20 percent ad valorem) is well known; in some cases these duties may represent an effective *de facto* devaluation on the import side. This tariff wall is augmented in a number of countries by quantitative import restrictions and special fiscal charges. Yet on top of these sheltering restrictions, domestic cost conditions seem to have dictated a substantial added measure of protection for many local products, of which perhaps the most significant are cotton textiles.

The position of cotton textiles is worth special attention because their survival (or further development) apparently still requires protection well above the average level of revenue duties even in countries where the size of the domestic market is not a limiting factor in obtaining economies of scale or full utilization of capacity. The subject deserves much closer investigation than it has apparently received. Available information suggests that high labor costs, reflecting a number of features of the African labor market, are primarily responsible, but an overvalued exchange rate cannot entirely be ruled out, nor can some export subsidies in some competing countries. We have here, apparently, largely an "infant industry" situation, but the persistence of a noncompetitive performance in several countries seems to dramatize the general problem of infant industry productivity in manufacturing under African conditions, even where highly modern equipment is used. It dramatizes also another aspect of the trade problems of the developing countries; much of the low-cost textile production that African countries find it necessary to protect themselves against originates in India.

There are also reported cases of a high level of labor productivity in specific African manufacturing and mining operations.<sup>5</sup> Obviously much depends on the relative requirements of particular industries for unskilled, semiskilled, and supervisory labor. Another consideration is that such enterprises frequently have a large proportion of expatriate managerial and supervisory personnel. It cannot be inferred that a rapid expansion of such units would maintain in the short run, the same average level of productivity. These points may be subsumed under the category of infant-industry problems, but the real burden of bearing the costs in the short run and overcoming them in the longer run must be reckoned with.

As industrial production reaches out from the limited list of consum-

<sup>5</sup> See, for example, H. Ward, "Economic Spotlight on Nigeria," *SAIS Rev.*, Spring, 1963 (School for Advanced International Studies, Washington, D.C.), p. 23.

ers' and producers' goods mentioned above, the constraints imposed by the size of Africa's Balkanized markets became heavier and the question of economies of scale becomes crucial. Leaving aside one or two large countries, there are now about fifty countries with an average money income not much greater than that of a Western European city of 100,000. With due allowance for differences from European consumption patterns, little industrial expansion, particularly in heavy industry and durable consumers' goods, can be based on markets of this size, even if other obstacles affecting productivity do not intervene.

It may be unnecessary to emphasize that little increase in effective demand can be expected merely from the creation of economic unions in Africa by the stroke of a pen. A great deal of infrastructural investment—in transport, communications, power generation, and marketing facilities—will have to precede or accompany any significant expansion of intra-African trade; the process will also be facilitated by some changes in monetary systems. The less dramatic expansion of intra-African trade in foodstuffs (which is probably now underestimated in the statistics, though still not much more than 4 percent of total African exports) may offer greater returns in the near future. But even the growth of this trade requires considerable infrastructural preparation.

The problem of arranging an equitable distribution of the costs and benefits of regional or subregional economic integration must also be reckoned with, particularly when, as in the African case, a substantial amount of protection against overseas imports may be required for a protracted infant industry period. Pressures exist for disintegration of existing subregional markets, as well as for integration; Northern Rhodesia's reported pleasure at the prospect of being able to import Indian textiles at a price considerably below that required by Southern Rhodesia's mills following the impending dis federation cannot be interpreted merely as a political expression.

In some cases, furthermore, integration of the national economy will have to precede or accompany a wider integration. The sooner a beginning is made on all these things the better; recognition that the road is long and that certain subregional groupings may have a higher degree of complementarity than others should help to smooth the path. It will be important, also, to maintain and hopefully increase the already substantial amount of intraregional labor mobility. And the process will be facilitated by any technological advances that permit reduction in the optimum size of plant.

Despite its recognition of the dynamic aspects of comparative advantage, can it be said that the foregoing "model" lacks sophistication? Where, it may be asked, is the underemployed labor, and where the shadow prices?



Expansion in three broad directions still remains productive for most, if not all, African countries: exports of primary commodities overseas, production of food for domestic or intra-African consumption and manufacturing for the African market. If, as seems reasonable, some increase in agricultural productivity in export crops is assumed, export expansion and import-substitution are not exclusive alternatives.

The first point to be made is that expansion in each direction will require considerable investment (including investment in the training of labor at almost all grades of skill, including managerial skills); a comparison of the "profitability" of allocating scarce capital resources in each of these directions will be warranted, and differences in factor endowments and stage of development will lead to different conclusions in particular countries. Second, with due allowance for a few countries in which Malthusian pressure suggests an imputed cost at least for unskilled labor of zero or thereabouts, the general position in tropical Africa suggests the existence of some measure of competition among the three sectors for labor of any significant degree of skill. The opportunity cost of labor in import-substitution is not necessarily zero, although the possibility of relatively cheap increases in agricultural productivity through improved techniques should reduce the cost.

To these considerations may be added a third; namely, the effect on income distribution of local high-cost production of items of mass consumption formerly imported, such as cotton textiles. It may be true that in one way or another any internal savings for development must be squeezed at this stage largely from the agricultural sector; but the claims of the present generation of consumers must be weighed and appropriate fiscal action taken to correct inequities.

#### IV. *Conclusion*

Whether much has been added to economic theory by the recent emphasis on the dynamic aspects of comparative advantage may remain a subject of debate. There can be little doubt, however, that the sharper focus on the dynamic aspects has had a healthy impact on considerations of policy in developing countries. For one thing, there is a greater awareness of the importance of taking the export sector explicitly into account in development programs. For another, there is a heightened concern with the consequences that may be visited upon economies lacking a capacity to transform their structures in the face of changing world markets.

As the last continent to enter the world economic scene, Africa has been in a position in which the real cost of fostering many industries in a sheltered domestic market was a formidable one in comparison with exploiting the path of export expansion. Both agriculture and

# IMPORT SUBSTITUTION AND INDUSTRIALIZATION IN BRAZIL<sup>1</sup>

By WERNER BAER *and* ISAAC KERSTENETZKY  
*Yale University*

We shall analyze some aspects of the industrialization policies followed by the Brazilian government in the post-World War II period. We shall concentrate on the effects upon the structure of the economy, the rate of growth, and the performance of the principal new industries.

Let us state what we understand to be the realms of discourse comparative cost and development theorists are engaged in. The former are concerned principally with questions related to the structure of international trade and the relative efficiencies of those industries in which various countries are specialized. The latter's concern is chiefly with the maximization of the rate of growth and their analysis runs mainly in terms of the rate of investment, changes in the economy's structure, and "linkages" of industries in which investments were at first concentrated. The former is concerned with the maximization of world production, consumption, and trade in a static setting, while the latter is concerned with the maximization of growth of a limited area.

## *Inevitability of Industrialization*

Brazil's industrialization should be viewed against a background of declining earnings of its traditional exports, which consist principally of coffee, cocoa, sugar, and cotton. From the long-term trends in world trade, it would seem that these commodities are not among those which have a bright future. A glance at Table 1 should make it clear that primary goods, in general, and food and agricultural raw materials, in particular, have steadily lost in relative importance in world trade. Part B) of the table indicates that world imports and the imports of industrial countries from nonindustrial countries have been shrinking considerably, much of this shrinkage being due to the decline of the relative share of Latin America.<sup>2</sup>

Further evidence of the dim outlook for the exports of primary producing countries and their continued reliance on specialization in those goods has recently been shown by the U. N. *World Economic Survey, 1962*. The following estimates were obtained for the income

<sup>1</sup> The analysis of this paper will be concerned mainly with events up to and including 1961. We believe that the economic difficulties since then are due to the political crises and not to the past industrialization policies.

<sup>2</sup> This decline would have been even greater had petroleum and petroleum products been excluded.

TABLE 1  
CHANGES IN THE STRUCTURE OF WORLD TRADE

A. *World Exports of Merchandise*  
(Percentage Distribution at Current Prices)

	World			World Excluding Iron Curtain Countries	
	1913	1929	1937	1913	1953
Food.....	29.0	26.1	24.8	27.0	22.6
Agricultural raw materials.....	21.1	20.0	19.5	20.7	13.9
Minerals.....	14.0	15.8	19.5	14.7	19.8
Manufactures.....	35.9	38.1	36.2	37.6	43.7
	100.0	100.0	100.0	100.0	100.0

SOURCE: Lamartine P. Yates, *Forty Years of Foreign Trade* (London: George Allen & Unwin, Ltd., 1959).

	1948	1953	1958
Primary goods.....	55.5	51.0	48.2
Manufactured goods.....	44.5	49.0	51.8
	100.0	100.0	100.0

SOURCE: Joseph D. Coppock, *International Economic Instability: The Experience After World War II* (McGraw-Hill Book Co., 1962).

B. *World Imports by Geographical Areas*  
(Percentage Distribution)

Imports from→ to ↓	Nonindustrial Areas			Latin America		
	1953	1960	1961	1953	1960	1961
<i>Industrial Areas</i> ..... (excluding Eastern Europe, including Japan)	37.4	28.3	27.1	12.9	8.7	8.0
<i>World</i> .....	31.5	24.8	24.3	9.8	6.8	6.5

SOURCE: GATT, *International Trade* (1961).

elasticity of imports of the industrially advanced countries from the developing countries.<sup>3</sup>

<sup>3</sup> "These estimates were derived from regression of gross domestic product of the industrially developed countries on imports of each commodity group from the developing countries. The sample covers the period 1953-1960." U.N. *World Economic Survey, 1962*, p. 1. "The Developing Countries in World Trade," p. 6. In the case of coffee, Brazil's chief export product, a study has shown the existence of low price and income elasticity in the U.S. It was found that an increase of 10 per cent in the price of coffee resulted in a 2.5 per cent reduction of consumption, while an increase of 10 per cent in real consumers' income resulted in a 2.5 per cent increase in coffee consumption (see Rex F. Daly, "Coffee Consumption and Prices in the United States," *Agric. Econ. Res.*, July, 1958).

<i>Commodity Group</i>	<i>Income Elasticity</i>
Foodstuffs (SITC groups 0 and 1).....	0.76
Agricultural raw materials and ores (SITC groups 2 and 4).....	0.60
Fuels (SITC group 3).....	(1.40)
Manufactured goods (SITC groups 5 to 8).....	1.24

It is also well known that consumption of raw materials by the industries of industrial countries tends to increase at a slower rate than production due to more efficient techniques and the development of synthetics. It is in this context that one should view the industrialization of Brazil.

Although Brazil came out of the second World War with substantial amounts of foreign exchange reserves, these vanished within one year in an import spurt, a large proportion of which consisted of consumer goods. Thereafter, direct controls were instituted, mainly exchange controls, which continued in one form or another until the present time. Even though the terms of trade became increasingly favorable up to 1954, the import needs of the country increased to such an extent, due to the government's development aims, that controls had to continue. And, despite the fact that the present terms of trade are still above those of the immediate postwar period, there has been a basic agreement among policy-makers that import-substituting industrialization was absolutely necessary to achieve a high rate of growth. The need for continued high rates of growth has become especially acute in the 1960's because of increasing population pressure. The population growth rate already increased from an average of 2.4 percent per year in the decade of the 1940's to over 3 percent in the 1950's.

### *Industrialization Policies*

Prior to the postwar era, Brazil did not pursue a systematic industrialization policy. The industrialization which did take place was always a by-product of external crises which limited the supply of imports, such as the two World Wars and the depression. The post-World War II industrialization spurt also began as a measure for coping with external difficulties rather than as a measure to actively protect and promote import-competing industries. The import licensing system from 1947 to 1953, the multiple exchange rate system from 1953 to 1957, and the modified system prevailing since that time only slowly became conscious instruments for the active promotion of an industrial complex.

Although critics at first claimed that these direct trade control measures offered protection to "nonessential" consumer goods industries rather than to industries of a more basic nature, a number of additional measures adopted in the 1950's, when policy-makers shifted the em-

phasis from mere balance-of-payments defense to active industrialization promotion, served to encourage a simultaneous development of the more basic industries.

In early 1955 a decree was issued by the central bank authorities (SUMOC<sup>4</sup> Instruction 113) which enabled foreign investors to import capital equipment without the need for exchange cover, if the investment was deemed desirable by the authorities for the development of the country. This was a great stimulus to the inflow of foreign capital, much of it directed into industries considered basic, like motor vehicles, steel, etc.

Brazil also made use of an old legal tool which was revised during the 1950's, the so-called "law of similars." Already before the first World War Brazilian manufacturers who were producing or intended to produce goods similar to the ones imported could apply for protection. In the 1950's the registration of a product as a "similar" became the basis for a substantial amount of tariff protection and for placing it in a high and protective exchange rate category.<sup>5</sup> The exact definition of a "sufficient quality and quantity" necessary to warrant the application of this law was left rather flexible, and it seems to have been applied in such a way as to encourage a substantial amount of vertical integration, either within firms or within the country by the emergence of supplying firms. Due to fear of outright exclusion from the market, foreign firms would often establish industries supplying the products needed by their initial plants, but the law also stimulated much local capital to establish supplying firms. It could thus be claimed that even if the protective devices used by the government stimulated industries of a nonessential nature, complementary policies provided substantial incentives for vertical integration and thus for the ultimate establishment of a "productive base."

The government, however, adopted further means to stimulate basic industries more directly. In the second World War it built a big steel complex at Volta Redonda with international financing. Also, in the 1950's a development bank was set up whose purpose was to finance certain infrastructure projects and certain key private and government enterprises in fields deemed essential by the government, such as iron and steel, chemicals, transport equipment, and machinery.

### *Effects of Industrialization Policies*

The high rate of real growth experienced by the Brazilian economy in the postwar period (see Table 2), and especially in the 1950's, can in

<sup>4</sup> Superintendency of Money and Credit.

<sup>5</sup> Lincoln Gordon and Engelbert L. Grommers, *United States Manufacturing Investment in Brazil: The Impact of Brazilian Government Policies 1946-1960* (Div. of Res., Grad. Sch. of Bus. Admin., Harvard Univ., 1962).

TABLE 2  
INDICATORS OF BRAZIL'S GROWTH AND FOREIGN POSITION

Fixed Investments GNP (percentages)		Rate of Real Growth	Terms of Trade (1953=100)	Export Quantum (1953=100)	Balance of Payments: Current Account Balance (Millions of U.S. \$)	Foreign Direct Investment (Millions of U.S. \$)
1947.....	17	1.8	45	127	n.a.	n.a.
1948.....	16	9.5	44	131	n.a.	n.a.
1949.....	15	5.6	53	117	n.a.	n.a.
1950.....	13	5.0	93	102	+104	28
1951.....	16	5.1	95	109	-470	70
1952.....	16	5.6	90	90	-709	118
1953.....	13	3.2	100	100	+ 17	109
1954.....	17	7.7	134	86	-235	75
1955.....	14	6.8	118	100	- 34	109
1956.....	13	1.9	113	108	+ 7	248
1957.....	13	6.9	117	100	-299	356
1958.....	14	6.6	119	96	-266	230
1959.....	16	7.3	109	117	-311	214
1960.....	15	6.3	101	118	-509	137
1961.....	n.a.	7.7	97	128	-241	169
1962.....	n.a.	4.1*	88*	114*	n.a.	n.a.

\* Preliminary Estimates.

SOURCES: Fundação Getúlio Vargas, *Conjuntura Econômica* and *Revista Brasileira de Economia*, Marco, 1962; SUMOC, *Boletim*, Maio, 1963.

large part be attributed to the industrialization policies. This is clearly indicated by the different sectoral growth rates. While the real product increased by 128 percent from 1947 to 1961, the real agricultural product increased by only 87 percent; the industrial product, however, increased by 262 percent. For the absolute increase of GNP between 1947 and 1961, agriculture was responsible for only 18 percent, while the nonagricultural sector contributed the rest. The key element here was the direct and indirect effects of the more than tripling of the industrial sector. All this is re-enforced when one considers the decline of earnings from agricultural exports during most of the decade of the 1950's. It should be noted that the fixed investment proportion was relatively low during the entire period under review, averaging 15 percent, which implies a low capital-output ratio. We shall have some more comments on this phenomenon in a later section of the paper.

Due to the high import content of investment, the investment proportion was correlated with the balance-of-payments deficit. Especially during the latter part of the period examined, the investment coefficient was maintained by a substantial inflow of private foreign capital.

One indication of the transformation of the economy resulting from these developments is the change in the income distribution by branches

of activity. As measured at 1947 constant prices, the share of agriculture in the net domestic product declined from 27 percent in 1947 to 22 percent in 1961, while industry increased from 21 percent to 34 percent in the same period.

An examination of changes in the structure of the manufacturing sector must, of course, begin with a brief review of changes in the import structure. One important fact which should not be overlooked is the downward trend in the ratio of imports of goods and services to gross domestic product. A first glance at Table 3, which shows the changes in the commodity structure of imports, reveals an important decline in the share of processed goods from 86 to 68 percent between 1949 and 1962. A large share of the increased proportion of raw materials imported represents goods not available in sufficient quantities in Brazil, like crude petroleum, which were, however, of extreme importance to the new industries and for the continued overall growth of the economy.

The newly protected industries did not only represent activities in the last stages of the production process. As can be seen in Table 3, and as will be shown further on with different types of evidence, the newly emerging industrial structure was fairly well balanced both from a horizontal and a vertical point of view. Important items of import substitution can be noticed both from a decline in their share of total imports and in their decline in real terms in relation to the average of 1949-50 (see column 3). Important substitution also took place in items whose share did not change or even went up, and whose real amounts of imports rose because these increases were all substantially less than the increase of industrial production which more than tripled. (This is re-enforced when we realize that for only three categories imports rose by more than real GNP, which doubled in the period.)

Table 4 shows changes in the structure of Brazil's industry according to the proportional distribution by gross value added and by workers employed between 1950 and 1960. It can be seen that the traditional industries of textiles, food products, and clothing have suffered declines in relative position, while the most pronounced growth took place in such key import-substituting industries as transport equipment, machinery, electric machinery and appliances, and chemicals. It is interesting to note that for the traditional industries there has been a greater relative decline in gross value added than in workers employed, while for many new industries the increase in gross value added was greater than the increase of workers employed. This is a reflection of the greater labor intensity of the more traditional industries as compared to the newer ones.

TABLE 3  
CHANGES IN BRAZIL'S COMMODITY IMPORT STRUCTURE\*  
(Percentage Distribution)

Commodity Groups	1949-50† (At Current Prices)	1962	Percentage Change of Imports in Constant 1949 U.S. Dollars Between 1949 and 1962
<i>Nonmetallic mineral products</i> .....	2.1	1.3	-1
Cement.....	1.1	—	-99
<i>Basic metal industries and metal products</i> .....	11.2	11.6	+61
Iron and steel.....	4.2	3.4	+27
Nonferrous metals.....	3.0	4.0	+108
<i>Machinery</i> .....	18.3	13.0	+6
Metal working machinery.....	7.0	7.6	+63
Other machinery.....	11.3	5.4	-29
<i>Electric machinery and appliances</i> .....	6.5	6.3	+46
<i>Transport equipment</i> .....	14.5	10.2	+5
Motor vehicles.....	9.1	2.4	-60
Other transport equipment.....	5.4	7.8	+112
<i>Paper and paper products</i> .....	2.6	2.6	+55
<i>Chemicals and products</i> .....	21.8	18.0	-0.3
Chemicals (proper).....	19.7	17.1	+5
(Products of petroleum and coal).....	(11.9)	(6.9)	(-43)
(Fertilizers).....	(2.1)	(1.2)	(-14)
Medicinal and pharmaceutical preparations	2.1	0.8	-42
<i>Textiles</i> .....	3.5	0.1	-89
<i>Processed food products</i> .....	1.9	3.0	+245
<i>Beverages</i> .....	0.4	0.2	+25
<i>Printing and publishing</i> .....	0.3	0.5	+320
<i>Miscellaneous</i> .....	2.5	1.2	-34
<i>Nonprocessed raw materials</i> .....	13.6	32.2	+45
<i>Total</i> .....	100.0	100.0	
<i>Change in industrial production</i> .....			+213
<i>Change in real GNP</i> .....			+105

\* The original data used were those expressed in current dollars.

† Average.

SOURCES: Serviço de Estatística Econômica e Financeira, *Comércio Exterior do Brasil*, several years. The basic data from this source were retabulated in order to make them comparable to the industrial census classification.



TABLE 4  
SECTORAL DISTRIBUTION OF GROSS VALUE ADDED AND EMPLOYMENT IN BRAZIL  
(Percentages)

	Gross Value Added		Employment	
	1950	1960	1950	1960
Nonmetallic minerals . . . . .	7.2	6.7	9.7	9.7
Iron and steel and metal products . . . . .	9.4	11.9	7.9	10.2
Machinery . . . . .	2.1	3.5	1.9	3.3
Electrical machinery and appliances . . . . .	1.6	3.9	1.1	3.0
Transport equipment . . . . .	2.2	7.5	1.3	4.3
Wood and wood products . . . . .	4.2	3.2	4.9	5.0
Furniture . . . . .	2.2	2.2	2.8	3.6
Paper and paper products . . . . .	2.2	3.0	1.9	2.4
Rubber and products . . . . .	1.9	2.3	.8	1.0
Leather and products . . . . .	1.3	1.1	1.5	1.5
Chemicals . . . . .	5.3	8.7	3.7	4.1
Pharmaceuticals . . . . .	2.8	2.5	1.1	.9
Perfumes, soap, candles . . . . .	1.6	1.4	.8	.7
Plastic products . . . . .	.3	.8	.2	.5
Textiles . . . . .	19.6	12.0	27.4	20.6
Clothing, shoes, etc. . . . .	4.2	3.6	5.6	5.8
Food products . . . . .	20.5	16.9	18.5	15.3
Beverages . . . . .	4.4	2.9	2.9	2.1
Tobacco . . . . .	1.4	1.3	1.3	.9
Printing and publishing . . . . .	4.0	3.0	3.0	3.0
Miscellaneous . . . . .	1.6	1.6	1.7	2.1
	100.0	100.0	100.0	100.0

SOURCE: IBGE, Recenseamento Geral do Brasil, 1960, *Censo Industrial*.

### *Measurement of Repercussions*

To the development economist the important criteria of success of a development program based on import-substitution industrialization are the direct and indirect impacts which such a program will have.<sup>6</sup> In Table 5 we tried to measure the repercussion effects which resulted from the industrialization of Brazil in the 1950's.

Our approach was as follows: we classified all imports according to the industrial classification used by the Brazilian Industrial Census. We then chose those industry groups where there was the greatest amount of import substitution between 1949-50 and 1962, as measured by the percentage decline or increase in the constant (1949-50) dollar value of imports of each category. (A more direct measure of import substitution such as the ratio of imports to value of output plus imports was not used because as this paper is being written, the pertinent data of the 1960 Census are not yet available.)

Next, we calculated for the census years 1949 and 1959 the percent

<sup>6</sup> In developing this section we were influenced by the ideas of A. O. Hirschman, *The Strategy of Economic Development*, and P. N. Rasmussen, *Studies in Inter-Sectoral Relations* (Amsterdam: North-Holland Pub. Co., 1956).

TABLE 5  
MEASURES OF DIRECT AND INDIRECT REPERCUSSIONS OF BRAZIL'S IMPORT SUBSTITUTION INDUSTRIALIZATION

	CHANGE OF SHARE OF TOTAL IMPORTS BETWEEN 1949-50 and 1962 (CONSTANT 1949-50 PRICES)	R*	CHANGE OF PERCENTAGE SHARE OF TOTAL BETWEEN 1949 AND 1959				BACKWARD LINKAGE	FORWARD LINKAGE	R*	RANKING OF TOTAL REPERCUSSIONS	
			Workers	R*	Gross Value Added	R*				A†	B†
1. Nonmetallic mineral products <i>Basic metal industries and metal products</i>	-0.8	7.5	0.0	11	-0.5	11	0.89	0.78	11	12	12.5
2. Iron and steel.....	-0.8	7.5	1.3	3	2.2	4	1.13	2.32	1	2.5	2
3. Nonferrous metals.....	1.0	11	0.1	10	0.2	9	1.17	1.65	2	8	7
4. Others.....	0.0	9.5	1.0	4	0.1	10	0.94	0.70	11	9	8.5
<i>Machinery</i>											
5. Metal working machinery....	-0.4	5.5	0.8	5	0.6	7	0.95	0.58	13	7	8.5
6. Other machinery.....	-6.1	3	0.6	6.5	0.7	6	1.07	0.81	8	5	6
7. Electrical machinery and appliances.....	-0.4	5.5	1.9	2	2.3	3	1.06	0.67	9	4	3
<i>Transport equipment</i>											
8. Motor vehicles.....	-6.7	2	2.3	1	4.9	1	1.30	1.01	5	1	1
9. Other transport equipment....	2.1	13	0.6	6.5	0.4	8	0.98	0.61	11	11	10
10. Paper and paper products....	0.0	9.5	0.5	8	0.8	5	1.04	1.68	3.5	6	5
11. Chemicals and petroleum and coal products.....	-7.8	1	0.4	9	3.4	2	1.10	1.59	3.5	2.5	4
12. Textiles.....	-3.4	4	-6.8	13	-7.6	13	1.03	1.13	7	10	12.5
13. Food, beverages, and tobacco.	1.1	12	-3.2	12	-3.6	12	1.27	0.91	6	13	11

\* Ranking (ranking of ties were taken as corresponding to the average of ranks which they jointly occupy).

† A—includes import substitution ranking. B—excludes the latter.

Sources: Column 1 computed from data in *Comercio Exterior do Brasil*, columns 2 and 3 computed from IBGE, *Censo Industrial*, Brasil, 1950 and 1960; columns 4 and 5 computed from W. D. Evans, and Hoffenberg, "The Interindustry Relations Study for 1947," *Rev. of Econ. and Statis.*, May 1959.

of the gross value added and total workers employed for each of these industries.

Using data from the 1947 input-output table of the United States, we computed the indexes of dispersion (backward linkage) and sensitivity to dispersion (forward linkage), using the following formulas:

$$U_j = \frac{\frac{1}{m} Z_j}{\frac{1}{m^2} \sum_{j=1}^m Z_j} \quad (j = 1, 2 \dots m) \quad \text{and} \quad U_i = \frac{\frac{1}{m} Z_i}{\frac{1}{m^2} \sum_{i=1}^m Z_i} \quad (i = 1, 2, \dots m)$$

where  $U_j$  = index of the power of dispersion,  $U_i$  = index of the sensitivity to dispersion,  $Z_j$  = sum of the row elements of the transposed inverse matrix,  $Z_i$  = sum of the column elements of the transposed inverse matrix,  $m$  = number of industries.<sup>7</sup>

The index  $U_j$  indicating the extent of the expansion induced by industry  $j$  in the economy as a whole corresponds to an estimate of what Hirschman called the "backward linkage effect."  $U_i$  indicates the extent to which industry  $i$  is affected by an expansion of the economy at large and is an estimate of the "forward linkage effect."<sup>8</sup>

We also made the same calculations for the main traditional industry groups: textiles, food, beverages, and tobacco. Due to the type of policy encouraging vertical integration, which we described above, we should not be too far off when assuming that the repercussion coefficients will work themselves out in the direction indicated and will not be substantially weakened by leakages through increased imports of needed supplies.

All this is summarized in Table 5. In addition, we ranked each of these measures. The backward and forward linkages were combined in a single ranking, giving double weight to the backward linkages.<sup>9</sup> The two last columns combine these rankings, representing an overall ranking of total repercussions. The next to the last column contains changes in the percentage of total imports at 1949-50 prices. However, in the last column we excluded the import-substitution ranking. This did not significantly change the overall ranking.

<sup>7</sup> Rasmussen, *op. cit.*, Chap. 8.

<sup>8</sup> The use of the United States input-output table can be justified on two grounds. First, the differences of technical coefficients of the manufacturing sector between various countries do not in our judgment significantly change the ranking of repercussions. Second, the U.S. table is particularly useful for our purposes because of the relatively small dependence of the U.S. manufacturing sector on imported inputs; i.e., the set of technical coefficients are not affected by imported inputs.

<sup>9</sup> As Hirschman indicates, backward linkages are more important than the forward ones because "... forward linkage could never occur in pure form. It must always be accompanied by backward linkage, which is the result of the 'pressure of demand.' In other words, the existence or anticipation of demand is a condition for forward linkage effects to manifest themselves." Hirschman, *op. cit.*, pp. 116-17.

From the table we reach the following conclusions. Those industries which show higher import substitution are at the same time those which ranked higher in terms of total repercussions in the economy.

One of the most dramatic types of import substitution occurred in the motor vehicle industry and the repercussions in terms of increased employment and gross value added were higher than anywhere else. It should be noted that the overall linkage effects are also the highest. Another high import-substituted product, machinery, also had a strong direct impact in terms of employment and value added, and relatively weaker, though not ineffectual, linkage coefficients.

The highest import substitution occurred in the chemical industry, especially in petroleum products. Its final repercussion rankings were somewhat lower, however, because of its low employment effect. Iron and steel are relatively low in terms of import substitution. It reveals a substantial area for further import substitution; i.e., an increase of steel producing capacity. (Some of this capacity is already being expanded.) A similar analysis would hold for the paper and paper products industry.

We have listed for comparison purposes the two big traditional sectors: textile and food products. They obviously grew at a much slower rate in terms of increased employment and gross value added; hence they lost considerably in relative importance.

Finally, one can see in Table 5 that in capital goods industries, such as machinery and transport equipment, the forward linkage is lower than the backward linkage. This is due to the fact that investment is treated in the input-output table as a final demand sector.

The picture which emerges from these linkages, that is, from the simultaneous growth of industries which to a large extent are each other's customers, is that of a remarkably balanced growth. This is not to say that Brazil did not experience imbalances. The latter occurred in some areas: e.g., between the growth of industries and the lagging of certain infrastructure facilities (one of the most acute being the expansion of power supply capacity), between the requirements for trained manpower and the dearth of training facilities, or the imbalances between the various regions, between agricultural production and the requirements of the growing population, etc. But it remains a fact that many complementary industries grew up simultaneously and acted as self-re-enforcing factors. This was mainly due to the policies discussed above.

#### *Capital-Output Ratio and Sectoral Growth Rates*

We already noted that Brazil's growth was characterized by a low capital-output ratio. Let us briefly account for this. The overall incremental capital-output ratio (ICOR) is a weighted average of sec-

TABLE 6A  
RATES OF GROWTH AND INCREMENTAL CAPITAL-OUTPUT RATIOS

	GROWTH (PERCENTAGE PER ANNUM) OF					ICOR				
	Period	GDP	Agriculture	Mining and Manufacturing	Other	Period	Overall	Agriculture	Industry	
									Total	Manufacturing
Norway.....	1952-60	3.4	-0.1	3.8	4.0	Index* 1955 (power and utilities =100)	9.8	55	15	n.a.
Denmark.....	1952-60	3.8	1.4	4.3	4.1	Index* 1955 (power and utilities =100)	7.3			
United Kingdom.....	1952-60	2.7	2.2	3.0	2.6		6.5			
Belgium.....	1952-59	2.6	2.1	3.1	2.7		5.0			
United States.....	1952-60	2.6	(1.5)	(3.0)	n.a.		5.0			
France.....	1952-60	4.2	2.5	4.8	4.2		4.0			
Italy.....	1952-60	5.9	2.4	8.8	5.5		3.6			
Germany (Federal Republic).....	1952-60	7.2	1.9	8.9	6.6	Index* 1955 (power and utilities =100)	2.8	29	11	n.a.
Portugal.....	1953-60	4.9	1.0	6.4	5.9	1952-57	4.0	3.7	1.7	1.6
Austria.....	1952-60	5.8	2.7	7.0	5.8		3.8			
Venezuela.....	1952-60	7.2	4.7	7.8	7.0		2.1			
Turkey.....	1952-60	4.7	2.7	4.8	7.0	1950-58	1.5	0.6	1.7	1.7
Greece.....	1952-60	6.2	4.3	8.1	6.2		2.2			
Philippines.....	1952-60	5.7	2.9	9.3	7.0		1.0			
Burma.....	1954-60	4.6	3.0	15.5	5.5		3.0			
Colombia.....	1952-59	4.5	3.4	6.4	4.6	1945-53	3.7	4.5	2.6	2.8
Ecuador.....	1952-60	4.7	3.5	4.8	5.8	1950-55	2.3	1.7	2.7	3.1
Thailand.....	1952-60	5.2	4.1	4.3	6.4		2.8			
India.....	1952-60	6.0	4.7	9.9	3.5		2.3			
Brazil.....	1952-59	3.0	2.3	2.9	3.9		2.3			
Chile.....	1952-59	3.2	0.2	7.6	3.1	1950-55	4.2	0.9	2.6	n.a.
Argentina.....	1952-60	1.9	1.7	2.4	2.0		15.0	3.5	37.9	n.a.
Algeria.....	1952-57	8.8	3.3	6.4	11.3		2.5			
Malaya.....	1954-59	8.0	5.6	22.4	12.9		2.9			
Korea (Republic of).....	1954-60	4.8	4.9	19.4	5.7		2.6			
Nicaragua.....	1952-59	4.6	3.3	5.1	5.6		3.3			
Nigeria.....	1952-56	3.7	2.4	6.1	3.9		2.8			

\* Only sectoral ICOR's are presented as indexes.

Sources: U.N., *World Economic Survey 1959*, U.N., *Yearbook of National Accounts Statistics*; W. B. Reddaway, *The Development of the Indian Economy*; *Economic Report to the President*; *Revista Brasileira de Economia*.

TABLE 6B

SECTORAL DIFFERENCES IN CAPITAL-OUTPUT RATIO,  
INDEX OF INCREMENTAL CAPITAL-OUTPUT RATIO

(Median for Eight Countries; Index, Power and Utilities=100)\*

Power and utilities.....	100
Transport, communications, and storage.....	65
Agriculture.....	39
Industry.....	26
(Manufacturing).....	(24)
Trade and services.....	18

\* Canada, Denmark, West Germany, Japan, Italy, Netherlands, Norway, United Kingdom, United States.

SOURCE: U.N., *World Economic Survey*, 1959.

toral ICOR's. Thus, the level of the ICOR of the sector (sectors) which is (are) growing most rapidly will have a dominant influence on the overall ICOR. Unfortunately, for Brazil we do not have sectoral ICOR's. We believe, however, that the evidence from many other countries substantially documents our case.

In Table 6 we have listed for a number of countries the rate of growth of the gross domestic product and its components, the overall ICOR's, and, for those countries where available, the sectoral ICOR's. An examination of these data shows that the overall ICOR is strongly influenced by the ICOR of the most rapidly growing sector (or sectors).

For example, in the case of Greece, Ecuador, and India, although the ICOR in the industrial sector is higher than in agriculture, the overall ICOR is relatively low. This is because in those cases it is the agricultural sector which might be contributing to the lowering of the ICOR, since total output is growing faster than the industrial sector. In Colombia and Portugal, the ICOR of industry is smaller than that of agriculture, but the overall ICOR's are relatively high. Here the explanation lies in a fast growing tertiary sector with a high ICOR. In Argentina the overall ICOR is the highest of those listed; it is caused by an unusually high ICOR in the industrial sector due to unused capacity.<sup>10</sup>

Our conclusion for Brazil is that the low overall ICOR must be due to a low ICOR in the industrial sector, which is also the most rapidly growing sector. Besides the fact that the industrial ICOR is generally lower than that of the other sectors (see Tables 6A and B), the Brazilian industrial ICOR might also be influenced by a greater degree of labor

<sup>10</sup> "... In a succession of years dating from the late nineteen forties, industry operated well below capacity levels partly because the absence of growth in the fuel, electrical power and transport sectors, when combined with acute balance of payments difficulties, created severe shortages throughout the economy; thus the level of manufacturing production attained by 1954 did not exceed the level previously reached in 1948. At the same time, however, investment intended partly to overcome the sectoral bottlenecks depressing output in the economy as a whole continued to rise." U.N., *World Economic Survey*, 1959, p. 74.

intensity in Brazilian industries, especially in areas marginal to the firm, like the handling of materials.<sup>11</sup>

### *Concluding Remarks*

In his famous Wicksell Lecture, Nurkse said that:<sup>12</sup>

If in an underdeveloped country the stock of productive factors is growing, but if development through increased exports to the advanced industrial centers is for one reason or another retarded or blocked, there arises a possible need for promoting increases in output that are *diversified in accordance with domestic income elasticities of demand* so as to provide markets for each other locally, in contrast to output expansion for exports, which is *specialized in accordance with international comparative advantage*. That the increase in production for the home market in these circumstances must ultimately conform to the pattern of domestic demand expansion is indeed a platitude if not a tautology.

We have shown that due to shrinking world markets for traditional exports and due to its high development aims, Brazil found it necessary to engage in import-substitution industrialization. We have shown that in the postwar period, Brazil attained a high rate of growth with a relatively low savings ratio.

The industrialization proceeded on a broad front. It was not an industrialization promoting "final touch" activities, since the policies of the government forced a vertical integration. This made the process self-re-enforcing. It resulted in a great expansion of industries with high repercussion effects. The industrialization did not result, however, in a perfectly balanced growth, since lags occurred in such sectors as agriculture, infrastructure facilities, and there still exists little effort to diversify exports.

We do not believe that the positive elements of Brazil's experience are necessarily worth copying by all underdeveloped countries. Much of its success was due to the country's size and variety of resource availability. The Brazilian case is relevant either for countries of similar size or for a number of countries which have formed an economically integrated area.

It might now be asked whether Brazil's growth could not have been accelerated even more had it been more selective in its import substitution along comparative advantage. This might have resulted in a relatively greater expansion of traditional sectors and light industry. This course was not possible due to the barriers to entry in world markets for such products.<sup>13</sup>

We might generalize by saying that comparative advantage is not dead for choosing industrialization paths for developing countries.

<sup>11</sup> This will be treated more thoroughly in a forthcoming volume on the industrialization of Brazil.

<sup>12</sup> *Equilibrium and Growth in the World Economy: Economic Essays by Ragnar Nurkse*, edited by G. Haberler (Harvard Univ. Press, 1961), p. 317.

<sup>13</sup> The U.N., *World Economic Survey*, 1962, Part I, has an excellent documentation of the many barriers to the exports of developing countries to the markets of developed countries.

Traditional theory teaches us a division of labor according to comparative costs. In the past this implied a concentration of countries like Brazil on primary products. With the change in the world consumption structure, where primary materials play a decreasing role, a correct division of labor would imply a change in the industrial structure of advanced countries, which would stop producing some of the goods they at present produce. Since such a sacrifice has not been forthcoming in great enough quantity from the developed world, i.e., since developed countries have not changed their structure fast enough, according to comparative costs, underdeveloped countries were forced to adopt industrialization lines which went beyond the area which comparative advantage would normally have offered them. Thus, development policies and comparative advantage are not necessarily at odds with each other but require the adherence of both the center and the periphery to be valid.



## DISCUSSION

ALBERT O. HIRSCHMAN: It is now five years ago since Ragnar Nurkse called on us, in his Wicksell lectures, to focus, in the context of development, on "incremental" comparative advantage in lieu of the "established" comparative advantage of traditional international trade theory. For the young and rapidly growing body of knowledge termed economic development, five years are a long time; yet the papers before us hardly testify to much progress in giving concrete operational content to the concept of incremental comparative advantage. In discussing the growth and trade performance of three underdeveloped areas, our authors differ in mood, but not in conclusion: I believe they all tell us that no ways of identifying incremental as wholly distinct from established comparative advantage have as yet been discovered; only, while this finding leaves Messrs. Baer and Kerstenetzky quite unruffled, it clearly puzzles Mr. Chudson and rather depresses Mr. Malenbaum. This difference in mood reflects less, I surmise, the diverse personalities of our authors than the development profiles of the countries they deal with. In Brazil, industrialization based essentially on import substitution has been a big success, so we are told, and if new export flows have failed to materialize, this is wholly the fault of protection in the older industrial countries. In India, where the impact of industrialization on the economy as a whole has been smaller than in Brazil, the failure to develop new exports cannot be dismissed quite so easily since India pioneered among underdeveloped countries in exporting manufactures and, in discussions of new patterns of international trade, her potential comparative advantage in steel, for example, had been frequently and hopefully pointed out. In Africa, finally, the road to economic development via import-substituting industrialization is less appealing than it had seemed elsewhere a decade or a generation ago, both because of intrinsic characteristics of most of the African economies (small size and high costs even where economies of scale are unimportant) and because one would like to do better than some countries that have run into trouble precisely by following that road.

Why is it then that such difficulties are encountered in identifying "incremental comparative advantage"? How, in the first place, are we to define this term?

Nurkse probably used the term "incremental" (versus "established") to get away from the overworked dynamic-static antinomy; but in this way he made our task sound easier than it actually is. Clearly his incremental comparative advantage has nothing in common with the Graham-type analysis where some commodities always hover on the edge of comparative advantage and can be made to turn from importables into exportables at will, through tiny doses of deflation or devaluation. In the development context, incremental comparative advantage does not result from a marginal reallocation of given resources in response to relative price shifts, but is the outcome of a fundamental change in the structure of production and can

appear only after that change. Incremental comparative advantage thus requires a judgment about the kind of new economic activity a country is best able to undertake. Frequently such activities are identified by existing import flows and require initial protection to be built up so that we end up with the annoying paradox to which I have drawn attention some time ago that incremental comparative advantage may be revealed to a country through the pattern of its imports. Whether a country will be able ever to become an exporter of some of the items it imports and upon whose substitution it is only just engaging, is obviously difficult to foretell at that point and it seems a bit farfetched to plan for so distant and uncertain an eventuality. The temptation is then to limit one's horizon and active industrial planning to the domestic market, and to wait for the infant to grow up before sending it on missions around the globe.

But to give in to this temptation is perhaps a mistake. Competing on export markets is one of the most powerful spurs to efficiency and it is quite possible that industries which do not get exposed to this spur at a rather early stage in their growth will never achieve the point at which they can hope to compete successfully in world markets. For a variety of other reasons which it would be too long to detail here an early effort at securing export markets, if necessary with the help of subsidies, may therefore be indicated. We begin to have some examples of such efforts, either by inadvertence—the overbuilt Brazilian automobile industry is a case in point—or by design as with Venezuela's new integrated steel complex.

It would of course be a mistake to look for incremental comparative advantage exclusively among economic activities that result from the implantation of new "development blocs." We must at least examine with an open mind the possibility that the almost universal predilection for overvalued currencies among today's underdeveloped countries is holding back some export flows which might prosper, not as a result of "fundamental structural changes," but merely upon an adjustment in the exchange rate. Mr. Malenbaum intimates several times that export pessimism, presumably caused by poor export performance, has in turn made that performance even worse than it needed have been. No such doubts can be detected in the paper on Brazil even though the persistent lag of the exchange rate behind the domestic inflation certainly resulted in chronic overvaluation for the cruzeiro. The overvaluation had its uses since it permitted to tax the traditional agricultural export interests without actually saying or legislating so. But this underhanded way of squeezing the *fazendeiro* had a price: that of holding back exports that Brazil's vigorously expanding economy might otherwise have achieved, in spite of the existing trade barriers to which Baer and Kerstenetzky appeal as an explanation.

Malenbaum's critique of Indian development policies goes, however, much beyond a mere call for a slight change in the export-promotion-import-substitution mix. He appears to find that India has not yet discovered her true economic vocation which would consist in producing an output that faithfully reflects her factor endowments, "primarily for the domestic market, but also for international trade." Too bad that his paper

ends abruptly with this statement, for one would like to get at least a glimpse into his neo-Heckscher-Ohlin world.

Perhaps Chudson's paper is a help at this point. It has the merit of getting away from the now overly fashionable North-South dichotomy. The South—the underdeveloped or “third” world—is no more a monolith than either of the two camps in the North; in fact, it is shot through with quite a few conflicts of interest of its own. Thus Mr. Chudson shows how industrialization in Africa could turn out to be primarily harmful to Indian exports of cheap textiles, just as Africa's significant gains in the international coffee and sisal markets were secured at the expense of Latin-American and Indonesian producers, respectively. But these potential or actual conflicts of interest make it possible to perceive also openings for collaboration and joint ventures in what might be called export substitution. As Indian textiles are being replaced, arrangements could be made for the compensatory importation into Africa of more advanced Indian manufactures, such as, perhaps, textile machinery. Similarly Latin-American countries would be more inclined to letting Africans occupy a progressively larger share of the international coffee market if the African countries showed a real willingness to shop in Latin America for that continent's manufactures with the proceeds of their coffee sales.

Finally I wish to add a technical comment on the Baer-Kerstenetzky estimate of linkage effects for Brazil's industries. Nobody could be more pleased than I that an attempt at such an estimate is being made, and I thoroughly enjoy the statistical ingenuity and enthusiasm of the Baer-Kerstenetzky paper. But here and there their enthusiasm has carried them perhaps a little far. Suppose that one goes along with the use of the U. S. input-output table for 1947 as an acceptable substitute for a nonexistent Brazilian one; and, more reluctantly, with the combination of backward and forward linkages into a total “repercussions effect” through the device of attaching arbitrary weights to the two categories; then our intrepid authors would still have to tell us what meaning they are proposing to attach to the apparent association between high repercussions effect and a high degree of import substitution: Did Brazil concentrate its import-substitution effort on the industries with a high combined repercussions index by chance or by design? Is import substitution fated to do so or did Brazil enjoy some especially favorable circumstances? Or isn't this kind of relationship always to be expected when industrialization has taken the import-substitution path and has proceeded vigorously for a decade or two for the simple reason that it could not have been sustained if it had not succeeded in building up an interrelated complex through a variety of linkages?

I hope the authors will address themselves to these questions as their work proceeds.

WOLFGANG F. STOLPER: All three papers agree that there is no conflict between comparative advantage and economic development, although Baer and Kerstenetzky leave the issue somewhat in doubt. None of the papers discuss the theoretical issues involved, and it is therefore hardly the place

of the discussant to open this subject very widely. Nevertheless, a few comments seem in order because it does not seem clear whether all four authors mean the same thing by comparative advantage.

The point to be made (and stressed by Malenbaum) is that growth involves allocation and, of course, the mobilization of resources. Now, as long as a country does not isolate itself completely from the rest of the world and as long as no one argues that all trade is or can be bad, allocation of resources and the mobilization must necessarily be influenced by these international relations. A fair interpretation of "comparative advantage" would therefore be simply: what can we say about the mobilization and allocation of resources in a particular country subject to the fact that the country is part of a developing world economy? The question is put in this manner explicitly or implicitly by Chenery in the article cited and by S. Chakravarty in his *Logic of Investment Planning*, and their answer that growth and comparative advantage are quite consistent is to be understood in this context. It is, however, not necessarily the context of Baer and Kerstenetzky. I base my interpretation on the stress which the first-mentioned authors put upon expansion of trade and changes in the terms of trade, unlike Malenbaum who consistently talks about whether the production of India is competitive and a reasonable allocation of resources, or unlike Chudson who deals with the African context.

Conflicts between development and international integration can arise only if it can be shown that as a result of trade the total of resources available to an economy is decreased and its allocation made less efficient. This is not easy to show.

There is perhaps another point worth making: countries are underdeveloped for a number of reasons. They may be poor because they lack resources and the necessary know-how. Characteristically they have an incomplete industrial structure: certain essential complementary factors needed to create employment and increase output are missing. The composition of imports becomes, therefore, of crucial importance in addition to their volume. Finally, we do have, in reality, economies of scale and transport cost. This means, for the problem on hand, that demand must reach a certain critical minimum level before it becomes profitable to produce a good. An allocation of resources that maximizes growth will also maximize demand, which in turn will allow and induce import substitution without any deliberate policy in this direction, and it will even possibly induce the establishment of industries producing goods which have not been imported at all. It just will pay to establish an industry.

The fact that there has been import substitution is therefore quite consistent with allocation according to comparative advantage, and it is precisely what ought to be expected. If the theory of international trade is expanded to include location theory, the argument becomes, if anything, even more persuasive. Moreover, operationally speaking, the only way in which we can ever find out whether a country has a comparative advantage is the same as how one finds out whether import substitution is sensible or whether to invest at all: one just has to make an economic pay-

off calculation that takes alternative uses of factors and economic repercussions throughout the economy into consideration. If a country can produce competitively, it will do so; if not, it is highly dubious whether net output has increased as well as whether import substitution has taken place.

There is just one more theoretical comment I should like to make. Much of the theoretical discussion on trade and growth proceeds on the assumption that there are exportables and importables. However valuable this approach in itself, it ignores that there are goods which are neither; that with growth the composition of both exportables and importables changes, becomes more complicated, and indeed that individual commodities may switch categories. Both facts indicate that as growth proceeds there ought to be an increasing amount of domestic production of goods previously imported, or of goods not imported (because of transport cost) and not domestically produced; as the composition of trade changes, terms of trade statistics and indices of the quantum of imports and exports become increasingly difficult to interpret. Professor Malenbaum's discussion of India stresses the failures of planning to increase for various reasons, the exports of new goods. He correctly, in my opinion, feels that the production of new goods for exports is not as hopeless as the usual discussion based on a two-commodity model, or on the inelasticity of demand for traditional exports makes one expect. There is the case of Japan, whose very success seems quite irrationally to be considered proof that other countries could not do the same. To give another example: In Nigeria, Dunlop and Michelin are beginning to produce rubber tires and they plan to export as well as to satisfy the domestic market.

I should now like to turn to the Brazilian paper as the most controversial. The difficulties the reader has concern the facts of the case and the statistical procedures. Only secondarily are the theoretical issues, on which I have already touched, involved.

First, questions of fact. The Brazilian economy has grown substantially. It has also experienced a very severe inflation. This raises serious questions on the manner of deflation and of the measurement of the real growth of the economy as a whole and of the individual sectors. The paper points out that industry has grown more substantially than agriculture, which need not be doubted. Yet, some uneasiness remains, and it is not alleviated by the fact that successive issues of the U. N. *Yearbook of National Accounts Statistics* (the only source available to me) give quite different measures of real sectoral growth, using 1948, 1949, 1955-57 factor cost and market prices in a manner which makes comparison rather hazardous and cross-checking next to impossible.

If one pieces together the information on output of agriculture and industry in current factor cost, the growth of industry between 1950 and 1960 is about 20-25 percent greater than that of agriculture. Between 1955 and 1960, growth of agriculture in current factor cost was 311.6 percent, of manufacturing 344.6 percent; the real growth was, between 1955 and 1960, 118.6 percent for agriculture, 164.1 percent for industry—both based

on measures in market prices of 1949. This implies that agricultural prices, on the basis of 1955=100, have risen to 262.7 by 1960, and industrial prices to 210.0 percent. Moreover, throughout the whole period, the implied price rise for agriculture was greater than for industry. This should have made agriculture relatively more attractive than industry. It should also have been reflected in food prices. Yet, on the basis of 1955=100, the cost of living of all items in 1960 was 308.3 percent, for food alone 325.0 percent, but until 1959 the food index always rose slightly less than the cost of living index as a whole. Undoubtedly there is a rational explanation. But one feels doubts and that considerable more work is needed before one feels that all the facts are explained; and one suspects that industrial output (which includes mining and construction) may not have risen quite as much as the figures claim.

Second, procedures. Here I have considerable misgivings. B. and K. take the U.S. input-output table lock, stock, and barrel and apply it to Brazil, even though only to establish rankings of industries. Now I am aware that the procedure of taking coefficients from one country and applying them to another has the sanction of Leontief himself, and I admit to having done something of the kind myself. Nevertheless, I feel that B. and K. have overdone it. The Brazilian industrial structure is very different from the American and lacks quite a few American industries. One should therefore expect domestic repercussions to be different, even with the same kind of technology. In the case in which I have constructed an input-output table (see Table 1) for an economy, the coefficients for all the industries that were characteristic for that particular economy were derived from data of that economy, and only additional coefficients were taken from another economy; my collaborator and I made reasonably certain that the technologies and the economics of the two economies were reasonably similar. I feel as

TABLE 1  
COMPARISON OF RANKINGS BY BAER AND KERSTENETZKY WITH TWO ALTERNATIVES

INDUSTRY	IMPORT SUBSTITUTION	B. AND K.	ALTERNATIVES	
			1	2
1. Nonmetallic mineral products.....	5	10	10	9
2. Iron and steel.....	5	3	3	3
3. Nonferrous metals.....	8	4	3	4
4. Others.....	7	9	8	8
5. Metal-working machinery.....	6	7	7	5
6. Other machinery.....	3	4	5	3
7. Electrical machinery.....	6	4	2	3
8. Motor vehicles.....	2	1	1	1
9. Other transport equipment.....	10	8	6	7
10. Paper and paper products.....	7	5	4	4
11. Chemicals, petroleum, coal products.....	1	2	2	2
12. Textiles.....	4	7	9	7
13. Food, beverages, and tobacco.....	9	6	6	6

Alternative 1: Forward and backward linkage counted only once, employment, value added.

Alternative 2: As above, but including import substitutions.

if I now know what Brazil would look like if it looked like the United States in 1947.

Moreover, other data supplied by B. and K. actually contradict any comfortable assertion that for technological reasons it is possible to apply the American coefficients to Brazil. For, later in the paper, B. and K. show that power per worker differs substantially industry by industry in the two countries, which they attribute to differences in the efficiency of auxiliary labor in the two countries. Maybe so, but the input-output coefficients know nothing about this. The figures strongly suggest that the repercussion of an increase in any industrial output on machinery inputs is seriously overstated in Brazil by using American coefficients, since in U.S. industry auxiliary technology is more mechanized.

Next, the authors, using the American input-output coefficients, compute rankings of backward and forward linkage, and in establishing the final index of total repercussions give backward linkage a double weight, quoting Hirschman's observation that "forward linkage could never occur in a pure form." However, this seems to me—possibly erroneously—hardly to justify that double counting. Since the indices are calculated from an input-output table, they seem to me already to have taken any indirect repercussions into consideration, the whole point of an input-output table being that it shows all direct and indirect repercussions. In addition, I am not sure that the interpretation of Hirschman given by B. and K. is the only possible one. Fortunately we can ask him directly. Moreover, any growth of any industry immediately means an increase in incomes and demand everywhere throughout the economy. Perhaps forward linkage also should be double counted.

Suppose you recalculate the ranking of total repercussions consisting of forward and backward linkage counted only once, value added and employment only, but without import substitution.<sup>1</sup> Suppose you plot this recalculated combined ranking against the ranking of import substitution. The new picture is sufficiently different to make one wonder. The position on the graph of chemicals, motor vehicles, iron and steel, nonmetallic minerals, and food are unchanged. But electrical machinery, which ranks sixth in import substitution, now ranks with chemicals in total repercussions; nonferrous metals which ranks eighth in import substitutions, now ranks with iron and steel in third place in total repercussions, and other transport equipment has moved from eighth to sixth place in total repercussions while ranking tenth in import substitution. (See Table 1.)

While, therefore, there still remains a relation between the ranking of total repercussions and of import substitutions, it is considerably weakened by a measurement of the rankings which seems to me at least as defensible as the one made by B. and K.

My final comments relate to the calculation and use of the incremental capital-output ratios. I have grave doubts on the *ex ante* usefulness of ICOR, though I consider a judicious use of an *ex post* calculation defensible.

<sup>1</sup> I do not quite understand B. and K.'s argument why import substitutions should be included with total repercussions, if these total repercussions are to be compared with import substitutions.

It is useful to remember that output and capital are measured by index numbers and in particular that the measurement of output involves, or ought to involve, a measure of profitability. Even if physical output grows, the index of output does not grow if inputs cost more than the value of output. Assume that all is well in the measurement of capital and output. A low ICOR indicates as much that the industries were profitable as that the composition of capital inputs among different industries was of a particular kind. What B. and K. have shown, provided their figures are accepted, is that the industries developed by Brazil were profitable.

Once again I find, however, the statistical procedures highly dubious and inconsistent with other facts supplied by the authors. I am disappointed that B. and K. have found it impossible to make direct estimates of the distribution of investments among major sectors, and that they have relied once more on American figures on installed capacity per worker in the U.S. and Brazil, which, if anything, are indicative of capital stock and not of the addition to that stock. Their explanation that the overall incremental capital output ratio in Brazil is due to a high percentage of investments in industry may be correct, but such evidence as the authors present is equally consistent with the fact that investment in agriculture was particularly productive. Their Table 2 indicates that 4 percent of local currency loans of the development bank between 1952 and 1962 was given to agriculture, whose output between 1952 and 1961—the last year for which I have any figures—increased 52 percent, while 36 percent of the loans went into basic industry whose output rose 137 percent between the same two years.

Now obviously there have been other investments into both agriculture and industry. It does not seem to me, however, that a ninefold investment in industry compared to agriculture-raising output less than threefold in industry compared to agriculture suggests particularly low incremental capital-output ratio in the former compared to the latter. Since another 45 percent on investments went into electricity, whose output measured in kwh. increased 237 percent, and only 15 percent into transport, I do not believe that B. and K.'s explanation can be correct. The figures suggest that 25 percent more investment in electricity compared to industry raised electricity production more than twice as much as industrial production! Rather the explanation of the facts—if they are really so—is more flattering to Brazil. Personally I feel—and I am developing this in my book—that high capital output ratios measured *ex post* are not so much the measure of a peculiar distribution of investments among industries with higher and lower ICOR's, but are simply the sign of bad investments which have negative payoffs.<sup>3</sup> B. and K. have proved that Brazilians have not wasted their resources—provided, of course, that the facts are really as stated.

I feel somewhat uneasy at criticizing so heavily what I feel to be basically a fine paper. Yet I feel very uneasy about the fact that virtually all major conclusions on the Brazilian economy are derived from data on the American economy. I do not doubt that Brazil has grown very much, and I

<sup>3</sup> The reference to excess capacity in the case of Argentina makes basically the same point.



have not the slightest doubt that the growth is consistent with international trade theory. In particular, I see no conflict whatsoever in development according to domestic income elasticities of demand and according to comparative advantage as I tried to indicate at the beginning of my comments. I would have felt a great deal easier if B. and K. had given us some Brazilian data and shown that Brazilian relations are in fact the same as in the U.S. My comments have shown, I hope, that there are some legitimate doubts on that score.

RAYMOND VERNON: A few years ago, a panel on the subject of comparative advantage theory and economic growth almost certainly would have struck sparks. Today, I suspect, some of the fight is beginning to go out of the issue. The papers we are discussing certainly seem a good deal more eclectic on the subject than, say, Viner's Brazil lectures of 1952 and the writings of Prebisch and Singer of the same period.

To be sure, the Baer-Kerstenetzky paper rattles the sabers a bit by staunchly insisting on the need for a deliberate policy of large-scale import substitution as indispensable for growth. It dismisses the question whether such a policy is a relatively efficient path to growth, as compared, say, with the promotion of exports; it simply insists that increased exports are not possible, at least not in the case of Brazil. The structure of demand in the advanced countries for primary products plus the allegedly restrictive commercial policies of the advanced countries with respect to secondary products are taken as sufficient reasons for the avoidance of such a line of strategy.

Nevertheless, the paper seems implicitly to accept the relevance of comparative cost considerations in framing growth strategy. One is led to suppose that Baer-Kerstenetzky might advise that Brazil concentrate more of its development efforts on exports if only it could be assured that the advanced countries would agree to play the game by accepting such exports.

The Malenbaum and Chudson papers are nowhere near so grudging in accepting the relevance of comparative costs to economic development. As I read these papers, both insist upon the need by one means or another to induce changes in the production possibilities of an underdeveloped economy, while acknowledging that comparative costs may play a major role in the selection of the changes to be induced.

What has happened, it seems to me, is that the development economists have shifted their position during the past decade in a salutary and constructive way. While gratified by the gains from import substitution (such as the Brazil case purportedly demonstrates), they have also been disconcerted by import-substitution programs that fail to fulfill all their promise (such as India seems to represent). The pressure on internal costs from overly-rapid import-substitution has proved disconcerting in some cases. So has the continued reliance of the new industries on imported intermediate materials and machinery. Accordingly, the development economists have felt the need to reconsider the possibility of increased exports, preferably increased exports of manufactured products, as a way out of an increasing bind. But which exports? In an effort to answer that question,

development theory seems to be rapidly rediscovering and reformulating some aspects of comparative cost theory.

As far as the international trade theorists are concerned, a similar conciliatory trend has been in evidence. There has been increasing acknowledgement of the relevance of various versions of the infant-industry argument. The case of declining internal costs and internal economies of scale has been elevated from the footnotes to the text of trade theory discussions. And the possibility of significant external economies, accompanied by a divergence between private and social cost, also has gained slow acceptance. A recognition of somewhat less than perfect internal mobility of the factors of production in underdeveloped areas is now generally acknowledged, as is the related possibility that pockets of underutilized labor may exist. Accordingly, few trade theorists would be unwilling to accept the validity of Gerald M. Meier's statement that "the gains from trade can facilitate development, but they cannot be a substitute for the developmental forces that must necessarily be created within the domestic economy."

The increasing eclecticism of both camps strikes me as a helpful development. But from some points of view, I would be sorry if controversy were to disappear too rapidly. Now that comparative cost doctrines have begun to acknowledge the relevance of problems of scale, it seems to me that the usefulness of this body of theory for explanatory, predictive, or prescriptive purposes, could be increased much more by consideration of two other groups of problems.

One group of problems centers on the role of information and ignorance in determining the pattern of exports and imports. Under comparative cost theory, the implicit assumption is that all production and trade possibilities are known to one or more of the infinite number of enterprises involved. With equilibrium exchange rates and with constant or increasing costs, appropriate market responses then occur.

These assumptions, restrictive though they may seem, are far from crippling with respect to certain kinds of economies. The assumptions serve well enough in an economy whose exports are largely staple raw materials, selling to a well-advertised market through established channels of trade. And they do not serve too badly as long as the production structure of an economy is not changing very rapidly. If structural changes are occurring rapidly, however, the assumption of adequate information begins to become a handicap. And if export opportunities in differentiated products or differentiated markets represent the wave of the underdeveloped country's future, then the issue of ignorance and information is elevated to one of critical importance in any explanation, prediction, or prescription of export behavior by such countries. A theory which disregards the fact that most traders and producers are likely to be largely ignorant of their opportunities can only have a very limited relevance.

The applicability of comparative advantage doctrine relates not only to issues of information and ignorance but also to problems of risk—risk in the sense of the anticipated variance about a mean expectation. As economists, we tend to assume that given price levels will generate given responses of producers and exporters. If risk enters the picture in some way, we try

whenever we can to convert the risk factor into an element of cost or profit.

The inadequacy of this approach stems from the fact that where a given level of risk is perceived by a group of entrepreneurs, they may be totally unwilling to act, irrespective of the average profit to be derived. The reason for their unwillingness may be quite rational; a run of losses, for instance, could put them out of the game. Accordingly, the variance about a mean expectation may condition behavior as much as the mean itself.

Once more, it is worth observing that the suppression of this problem need not be crippling in every setting. In a comparatively stable situation, where most contingencies can be foreseen or can be hedged, perceived risk may be of very little importance. But this hardly describes the situation facing the entrepreneurs in less developed areas, especially as they confront problems of international trade. The reduction of risk, therefore, may prove more important for such a group than the maximizing of yield. A systematic formulation of concepts in this area as they affect international trade would be a valuable addition to comparative cost theory.

# CANADIAN-AMERICAN ECONOMIC RELATIONS: EFFICIENCY AND INDEPENDENCE

## THE CANADIAN TARIFF AND THE EFFICIENCY OF THE CANADIAN ECONOMY

By H. C. EASTMAN  
*University of Toronto*

Two effects of tariff protection on economic efficiency in Canada must be distinguished. One is the distorting effect of protection on international specialization which arises from the alteration of the relative prices of goods. The other is the effect of protection on the organization and behavior of particular industries. It is possible to conceive of an efficient grape growing industry in Scotland in which firms minimize their costs of production given the input prices facing them, but which would disappear in the face of imports were it not protected by a tariff, because of natural disadvantages and, perhaps, because some of its inputs were themselves protected. This paper is concerned, not with comparative advantage, but with the effect of the Canadian tariff on the efficiency of existing Canadian industries; that is, with the extent to which firms minimize their costs within the limits of existing knowledge and at input prices existing in Canada.

The observable phenomenon in many Canadian industries is that the total output of plants is smaller than that of their foreign counterparts, that runs in the production of each line of products are shorter, and that the number of lines produced by a single plant are often greater. The popular hypothesis to explain these short runs is that they are caused by the small size of the Canadian market to which Canadian manufacturers are restricted by foreign tariffs. But this simplistic theory only explains short runs in industries in which there is but a single plant—a rare occurrence. It casts no light on the determination of the actual excessive number of plants in multiplant industries.

Professor Bain has explained how a low ratio between the total output of the industry and the output at the lowest point on a firm's long-run average cost curve tends to inhibit the entry of new plants.<sup>1</sup> This analysis is applicable to the effect of tariff protection.<sup>2</sup> Foreign tariffs reduce the size of the market, and, at the lowered ratio of total output to efficient plant size, firms are less inclined to enter an industry

<sup>1</sup> Joe S. Bain, *Barriers to New Competition* (Harvard Univ. Press, 1956).

<sup>2</sup> S. Stykolt and H. C. Eastman, "A Model for the Study of Protected Oligopolies," *Econ. J.*, Vol. LXX, No. 278.

with plants of large and efficient size or, if already established in the industry, they are less inclined to add to capacity by large increments. Rather, additions to capacity are smaller than the lowest cost size, firms preferring to suffer higher than minimum costs in order to avoid some of the depressing influence of added output on price. The lower the ratio of market to efficient plant size, the more are firms apt to charge higher prices to cover the costs of less efficiency. However, if barriers to entry are very high, because of important economies of scale, product differentiation, or some other factor, overcrowding tends to be less because of the inability of new plants to enter. The potential advantage for productivity of lower costs of production are lost, nevertheless, if the firms in the industry incur high selling costs rather than setting low prices or making high profits.

The high prices necessary to cover high costs of production and selling or high profit margins can usually only be charged behind the protection from foreign competition that is provided by the Canadian tariff. Thus, it is the Canadian tariff that permits high prices and, insofar as high prices are set, that encourages the entry of plants and firms even if these must enter at a scale too small to be efficient because the increment to total output made by an efficient unit would depress prices to an unprofitable level.

An attempt to discover the extent to which the model I have briefly outlined in fact explains the structure of Canadian manufacturing industry requires relevant definitions of industries and an acceptable way of estimating how costs vary with the size of best practice plants and firms in these industries. I will first raise a few points in relation to these questions and then briefly describe a sample of industries and indicate some of the results that stem from its analysis.

The first step is a definition of industries that is appropriate to the model. For this purpose, an industry consists of plants producing goods which have a common best practice technique of production. This is obviously a practical rather than a theoretical question in particular cases, but its solution requires judgment as to the extent to which differences in techniques are significant in relation to similarities and the appropriate classification may differ from that of the census, as well, of course, as from patterns of ownership. For instance, it is clear that, for this purpose, the production of liquid detergent and of solid detergent are separate industries, for the former is a simple mixing process, whereas the latter requires a good deal of knowledge and a drying tower with a large capacity.

The next problem is that of discovering information applicable to Canada on the economies of scale of plants or firms. One source of information is industry opinion in Canada. The danger here is that

businessmen tend to equate actual size and the lowest point on a long-run average cost curve, confusing the behavior of costs within scales that are actually open to them in terms of profit maximization in a market of a particular structure with the more abstract problem, to them irrelevant.

The second source is engineering studies. The use of these studies, which are primarily of United States origin, raises the question of the importance of differences in relative factor prices between Canada and the United States in the determination of best practice techniques. Annual labor income per employed worker in Canada is about 25 percent less than in the United States, whereas the rate of interest on long-term government bonds is usually about  $\frac{3}{4}$  of 1 percent higher. Thus the ratio of the rate of interest to the wage rate is roughly 60 percent higher in Canada. Does this mean that the best practice techniques in Canada are sufficiently different from those in the United States to render inapplicable to the former country long-term average cost curves derived from data relevant to the latter? Yet another indication of economies of scale and minima-optima is actual practice in the United States. It may be the cause of some surprise to Americans concerned with the efficiency of their own industries to see these used to some extent as an indication of the ideal, but it is a question of standpoint, and data on a more efficient structure may be of considerable value, even if the plants involved are not perfectly efficient. The relevance of actual United States practice, as of United States engineering studies, depends on the insignificance of different relative factor prices for presently built best practice plants, but another problem is introduced by time and technical change. Actual industries are composed of old and new plants and the rate of obsolescence is affected by differences in relative factor prices. Since the size of best practice plants or machines changes over time, the average size of plants or machines in an efficient industry differs both from that dictated by modern best practice techniques and, in a Canadian industry, from average plant or machine size in the United States.

The theoretical problems involved here are straightforward. The production function facing an industry is composed of isoquants reflecting the wide range of possible techniques that could be used to produce a certain output in the present state of scientific knowledge if sufficient investment in designing were made to develop these techniques. However, at any time, the techniques that have in fact been developed are restricted to the narrow range that is believed to be profitable by engineers and businessmen. Thus the techniques that are readily available are concentrated on the isoquants near the expansion path drawn with the aid of price lines whose slopes reflect the relative factor prices

facing the principal manufacturers. Other methods of production that are technically feasible remain undeveloped because of the cost of design and the unlikelihood that they would prove profitable.<sup>3</sup>

Over time, two things happen. First, the progress of science causes isoquants to shift towards the origin because increased knowledge usually makes it possible to produce goods with less of all inputs, even though the proportions in which they are used at constant factor prices may change. Second, relative factor prices change and the expansion path shifts for this reason as well. The change in factor prices appears to firms in a particular industry as independent of the progress of knowledge in that industry.

Relative factor prices in the Canadian economy are similar today to those that prevailed in the United States in the past and therefore in an anterior stage of technical knowledge. Three choices are open to firms in a Canadian industry. The first is to incur the fixed costs of designing methods based on present knowledge and appropriate to domestic relative factor prices. The second is to adopt the new United States techniques even though they are not exactly suited to domestic factor prices. The third is to adopt the old techniques of the United States which are appropriate to domestic relative factor prices, but based on a backward state of knowledge. The Canadian industry assumes the costs of designing new methods if this cost falls short of the costs of producing with inappropriate factor prices or with obsolete techniques. If the most suitable methods of production are not designed because of the high cost of designing, Canadian production costs exceed the level that would be attainable were the output of the industry larger. This cost of having a small total output is always positive, for the industry cannot escape some designing costs and shares the cost of developing the techniques suited to the United States industry in the price of machinery or in some other ways when it uses those techniques. The smaller the Canadian industry's output, the less is it likely to incur the fixed costs of designing, so that the use of techniques imperfectly suited to domestic conditions is one of the costs of small output.

Turning now to experience, it appears to be generally the case that the basic equipment used in manufacturing industries is the same in all industrialized countries. This is a recurring theme in international comparisons of manufacturing techniques. The minimum optimum size of plants in these countries is therefore the same for the same industry. Opportunities for varying the proportion in which factors are used arise where equipment is made especially for each installation as is the case, for instance, in petroleum refining or in cement and newsprint making. These are typically processes that are highly capital intensive

<sup>3</sup>W. E. G. Salter, *Productivity and Technical Change* (Cambridge Univ. Press, 1960).

even in low-wage countries. Relatively cheap labor does not affect the basic design because total labor costs are low. Hence techniques tend to be uniform internationally even in these cases. Exceptions leading to the adoption of outdated United States techniques in lower-wage countries occur only where technical progress in that industry has been slight, but the relative factor prices are markedly different. The relatively small size of most Canadian manufacturing industries and the small difference in factor prices between Canada and the United States compared to that between the United States and most other countries makes even more applicable to Canadian industry than to others the standards found in the United States.

The proximity of the United States to Canada and the high percentage of United States ownership of Canadian secondary manufacturing industry, which often entails some management trained and experienced in the United States, reduce the cost of transmission of techniques from the United States and therefore the cost of such techniques relative to techniques developed by special designing for Canada. These factors increase the likelihood that United States techniques are adopted in Canada by making them more economic, but also by making such adoption the easiest path for unenergetic managements.

Despite the rigidity that designing costs introduce in best practice methods, labor and capital are substituted for each other as a result of different relative factor prices. The point at which lower relative wage rates have their effect is in the greater use of labor in peripheral processes and in the rate at which equipment and plant are replaced. Though the basic design of equipment is not altered, certain processes, such as the handling of materials, are determined at the level of the plant or firm and tend to be less mechanized in Canada than in the United States. Relative factor prices also determine the moment at which existing plants and equipment are replaced. The rule is that equipment is obsolete when the variable cost of operating it is equal to the total cost of operating new equipment, and the point is affected by the prices of both labor and capital. A fall in either the price of equipment or the rate of interest reduces the total cost of operating new equipment and so hastens the replacement of existing equipment. Similarly, a fall in wages reduces the variable costs of existing equipment and increases the time for which it is used. Thus relative factor prices determine standards of obsolescence.

This discussion indicates that the substitution of factors in response to different relative prices occurs in the proportion of output produced by new plants and equipment and in the degree of mechanization of peripheral processes, and only exceptionally in the basic design of the main items of new equipment. The minimum optimum size of best



practice plants is therefore in most cases the same in Canada and the United States, but the average size of plant may differ between the two countries owing to a slower rate of obsolescence in Canada. Consequently the appropriate indication of the efficiency of plants in Canada is the size of recent replacement or of new additions to capacity.

The efficiency of Canadian industries is not solely a function of the efficiency of their plants because economies may arise from operating more than one plant in a single firm. Thus a distinction must be made between the average cost curve of the plant and that of the firm. The former is chiefly determined by economies of increased dimensions or of linked processes in equipment. But some costs, such as those undertaken for marketing or research, are external to the plant though internal to the firm and may decline over outputs greater than the minimum optimum size of plant. In that case there are economies of multiplant operations if the plant cost curve rises or if transportation costs limit the market for a single plant. What is more significant for an analysis of the structure of Canadian markets is that economies of multiplant operations may exist where the plants are inefficiently small whether or not these economies would exist if the plants were of minimum optimum size. The number of firms that operate in Canada depends on whether or not economies of multiplant operations exist when plants are of their actual size, their actual size being determined in turn by the pattern of competition in the national or regional market. The cause of the economies of multiplant firm operations in manufacturing becomes significant in explaining the type of ownership. If the economies of firm scale are chiefly owing to marketing relations, ownership is likely to be Canadian. The market to be influenced being in Canada, no advantage accrues to firms with foreign connections. However, if the economies of firm scale reside in the manufacture of technical knowledge, ownership is likely to be foreign. Foreign firms or firms affiliated with them entering only at small size because of the combined effect of foreign and the Canadian tariffs suffer only from diseconomies of plant size, whereas firms of independent Canadian ownership entering at the same size suffer from both diseconomies of plant size and diseconomies of firm size. In the cases in which economies of firm size are so significant that firms in Canada, whether with single or multiplant operations, do not exhaust them, Canadian firms, however established, tend to be bought out by foreigners.

A sample of industries in the manufacturing sector in Canada during the late 1950's was studied to determine whether the ratio of market size to the output of an efficient plant was a cause of efficiency or inefficiency in the Canadian market. The list of industries is the following: fruit canning, vegetable canning, cement, shipping containers, con-

tainer board, liquid detergent, solid detergent, newsprint, beef packing, pork packing, petroleum refining, basic steel, and tires. It thus comprises ten industries according to the Canadian census of industry classification and thirteen industries in terms of homogeneity of processes of production. It is the latter definition that is relevant for this study. The sample includes consumer and producer goods, goods with high and with low transportation costs in relation to their value, industries with national and industries with regional markets, industries with high ratios of market to minimum optimum plant size and some with low ratios, industries in which multiplant operations bring economies either at or below minimum optimum plant size and industries in which minimum optimum size is absolutely large and others in which it is small.

This type of analysis is obviously most easily applied to industries in which market size and minimum optimum sizes of plant and firm are most easily estimated. The validity of the concepts is independent of the ease of estimation, though it is certainly the case that the behavior of costs in multiproduct plants is difficult to ascertain. In a few industries, such as that producing heavy electrical apparatus, no consistent relation exists between size of plant and costs and these industries cannot be analyzed at all in this way. As it is, the results of a sample of only thirteen industries can in no sense prove anything, but give results that are suggestive of the effect of the tariff on industrial structure in Canada.

The efficiency of industries was measured by estimates of the proportion of individual pieces of equipment or of plants that were of minimum optimum size and over. The estimates of the proportion of capacity that was efficient was based where possible on the size of new installations alone.<sup>4</sup>

A positive relationship exists between the number of plants of minimum optimum size that could supply domestic production in Canada and the efficiency of existing Canadian capacity in that industry. The industries less than one-quarter of whose capacity was of the most efficient size were steel, fruit canning, tires, petroleum refining, and pork packing. For these industries the ratios of market size to minimum optimum plant size were  $2\frac{1}{2}$ , 4, 5, 7, and 15. The industries more than three-quarters of whose capacity was of the most efficient size were

<sup>4</sup>This is an excessively simple measure. It evaluates an industry entirely on the basis of the percentage of its capacity that is of minimum optimum size and over and gives no weight to the degree of efficiency of the remaining capacity, failing to distinguish between remaining capacity that is near minimum optimum size and capacity that is far too small, and between industries in which the long-run average cost curve of plants rises steeply at suboptimal scales and those in which it rises slowly. A more sophisticated measure of efficiency could be devised, but could not be used owing to inadequate information as to the behavior of costs at suboptimal scales in most industries.

solid detergent, cement, newsprint, beef packing, and liquid detergent. For these industries the ratios were 8, 18, 24, 42, and 49. The percentage of efficient capacity in shipping containers was estimated at 72 percent and the ratio at 27, while the remaining two industries, container board and vegetable canning, were efficient to the extent of 57 percent and 44 percent of their capacity with ratios of 5 and 21. These results are rough because they leave out of account the doubtless varying degrees of inefficiency of the inefficient capacity and because estimates of the proportion of capacity that was efficient were based only on the size of units. Nevertheless, it seems clear from the coefficient of correlation of .57r between the percentage of efficient capacity and the ratio of market to efficient plant size that a fundamental factor leading firms in Canadian industries to take excessive advantage of the protection offered by the Canadian tariff was the extent to which they would be interdependent if their plants were all of efficient size. In general, actual plants are less efficient; the fewer would be efficient plants in the market actually supplied. The advantage taken of the tariff is excessive in the sense that costs of production at efficient scale are below the costs actually incurred, not in the sense that any single firm could necessarily operate most profitably if it added to capacity units of efficient size. Businessmen in overcrowded industries presumably forecast correctly the consequences they evidently fear of large increments to capacity on the price of the product and on their profits.

Results about the efficiency of plants in regional markets are consistent with the results on a national basis. Transportation costs on the final products created more or less distinguishable regional submarkets in seven of the thirteen industries studied (fruit and vegetable canning, cement, shipping containers, container board, beef packing, and petroleum refining). The regional pattern of overcrowding was similar in each industry, with the larger market in Ontario and Quebec the least crowded. The smallest market, in the Atlantic Provinces, was the most crowded and those in the Prairies and in British Columbia were in an intermediate state. The unfailing association of smaller plants with smaller markets in this interregional comparison of the size of plants in the same industries confirms that the interindustry association of inefficiently small scale with low ratios of market size to minimum optimum size of plant is not accidental.

The efficiency of the operations of firms is not entirely revealed by the size of their plants in relation to the minimum optimum plant size because of the possibility of economies of multiplant operations. Thus the operations of firms may be less efficient than their plant size indicates if economies of multiplant operations exist and firms operate single plants. No such cases exist in the sample: all firms capable of

benefiting from multiplant operations having such operations whether their plants were mostly of efficient size (cement, solid detergent) or not (tires). On the other hand, firms may be more efficient than the suboptimal size of their plants indicates if economies of firm operations exist over an output greater than that of small plants and firms operate more than one plant. Wherever such economies exist in the sample, firms, in fact, operate more than one plant (petroleum, pork packing, vegetable canning). These economies of multiplant operations for plants of both optimal and suboptimal size can be usefully classified by their cause. Where the cause is marketing advantage, stemming perhaps from facing rivals in several product or input markets or from economies of national distribution, as it is in pork packing and cement or in petroleum refining, the firms typically operate more than one plant in Canada, and, in the former two industries, nowhere else. Where the economies stem from the production of knowledge, in the vast majority of cases firms operate other plants abroad (tires, solid detergent, vegetable canning, petroleum refining) and in some of these cases they operate only one plant in Canada (all tires, all solid detergent). This supports the view that foreign ownership is related to economies of firm scale that are external to the plant both where the plants are of optimal size and where they are of suboptimal size. However, the advantages of multiplant operation are greater the smaller the plant, for the smaller the plant the less it exhausts economies of firm scale within its own operations.

Multiplant firm operations also occur in the sample where economies of multiplant operations are nonexistent or at most very slight. Many causes for this may exist. In the sample, the chief reason is simply the growth of certain well-managed or fortunate firms. If this growth takes place in industries with regional markets (shipping containers) or with regionally distributed raw materials (newsprint, container board), large firms have several plants. Similarly, if growth takes place by the acquisition of other firms, such acquisitions leave a legacy of several plants to the large firm for a time at least (container board, fruit canning). Another cause of multiplant operations in one industry arises from the pattern of competition in another. In this sample, container board firms appear to restrict entry into that industry by buying up their customers, the more numerous shipping container manufacturers, thus creating multiplant operations in the latter industry. Finally, where some cost advantage from joint production exists, however slight, and one product has advantages of multiplant operations, this makes for multiplant operations for the other (beef and pork packing).<sup>5</sup>

<sup>5</sup>This latest point raises the question of joint production and the definition of an industry, a problem complicated by the fact that two products may be joint products at

Entry into an industry can be impeded, not only by the size of an efficient unit of production relative to total output, but also by the large amount of capital required to build a plant, the advantages that product differentiation gives established firms, and other barriers, especially legal restrictions and lack of access to knowledge. As does a tariff that impedes the entry of goods, barriers that impede the entry of new firms permit established firms to operate with higher prices and with higher costs or higher profits than would be possible in their absence. The entry into none of the industries in the sample is impeded by legal restrictions and lack of access to knowledge impedes the entry only of independent firms, not of plants belonging to multiplant firms. Since this analysis is of the efficiency of production and not of the nature of ownership of the firms, these two barriers are omitted from consideration. The two barriers relevant here are capital requirements and product differentiation and these vary within the sample from very high to low.

The capital required to establish a plant of minimum optimum size in Canada in the industries in the sample varied from \$100,000 for liquid detergent to \$200 million for petroleum refining and \$300 million for basic steel production. A rough negative relationship existed between the efficiency of the industries and capital requirements for which the coefficient of correlation is  $-.41r$ . The four industries in the sample for which capital requirements were approximately \$1 million or less were all efficient. Capital requirements were themselves inversely related to the number of efficient plants the Canadian market for a particular product could contain because plants that cost a lot also tended to be in industries with few plants. Nevertheless, capital requirements may have had an independent effect on industrial structure. Industries in the sample in which capital requirements for plants of minimum optimum scale were over \$20 million tended to have a lower ratio of minimum optimum plant size to market for a given level of efficiency than industries with low capital requirements. This is presumably because high capital requirements, being a barrier to the entry of new firms, permitted established firms to charge high prices without attracting new capacity while the Canadian tariff kept out imports.

It is the capital required to establish a plant of minimum optimum size, not a firm, that is important as a possible impediment to the entry

---

certain outputs and not at others. If the output of one or both goods is less than that of plants of optimal scale, the products may be joint products even if they are not joint products when output is sufficiently high to exhaust those economies of plant scale in processes applicable to either or both products that are the source of advantages of joint production at smaller outputs. In this sample, products were treated as those of separate industries where economies of joint production are unimportant at minimum optimum scale.

of new capacity whenever economies of multiplant operations are not caused by marketing advantages in Canada. However, in the cases in which advantages stem from multiplant production in Canada, the capital requirement is increased in proportion to the number of plants needed.

The efficiency of the industries in the sample appears to be unrelated also to different degrees of product differentiation ( $r = -.05$ ). In any case, the severity of the barrier erected by product differentiation to the entry of new firms is uncertain and may be slight, as is indicated by a comparison of the markets in liquid and in solid detergent. The markets for the two products were similar in most important respects including the efforts of established firms to differentiate their brands in the minds of the same group of customers. Yet new firms entered easily into the production of liquid detergent but were excluded from that of powdered detergent. Thus product differentiation was not an important determinant of market structure in the synthetic soap industry, notorious though it is for high promotional expenditures.

The conclusion suggested by an examination of this sample is that the principal determinant of the efficiency of manufacturing industries in Canada was the size of the market, measured in terms of the number of plants of minimum optimum size it could contain. Efficiency was also influenced to some slight extent by the size of capital requirements.

To turn now from efficiency to performance, it appears that the efficiency in production of the Canadian industry was, in turn, the principal determinant of the level of prices charged. For the purpose of this analysis, which is not concerned with international specialization but with the effect of the tariff on the organization of Canadian industries, the appropriate measure of whether a price is high or low is not a comparison with foreign prices but with the level of costs that it would be possible to achieve by efficient production given Canadian input prices. Costs, profits, and prices for each industry were rated as low, medium, or high. In eight of the thirteen industries (fruit and vegetable canning, newsprint, beef and pork packing, petroleum refining, steel, and tires) the rating of the prices and of the level of costs of production were the same, which implies that selling costs and profits were moderate in these cases, or that an abnormally high level in one was compensated by a low level in the other. In the remaining five cases (cement, shipping containers, container board, liquid and solid detergents) the prices set by the industry appeared to be high in relation to its costs of production, the difference between high prices and low or medium costs being due in three cases to high profits (cement, shipping containers, and container board) and in two to high selling costs (liquid and solid detergents).

In only two of the thirteen cases were low costs accompanied by low prices. One was the oligopolistic newsprint industry which took no advantage of Canadian protection because it exported. The other was beef packing, the market structure of which approximated perfect competition with an undifferentiated product that can be efficiently manufactured in small plants. Thus, except for these two cases, prices were higher than necessary to cover minimum costs of production and distribution and average profits so that either or all of costs of production, selling costs, and profits were high. A lower price for these products would either force the efficient organization of production, abolish excessive selling costs, or reduce profit margins in these industries.

The most readily available policy for lowering prices is the reduction of the Canadian tariff which, by allowing foreign suppliers to sell in Canada more cheaply, would oblige domestic producers to lower their prices. Such a policy would not be effective in all cases because other barriers to international competition than tariffs, chiefly high transportation costs, exist in some industries. Also, in certain instances, even with tariff protection the Canadian industry sets prices no higher than those of the foreign producers because of differences or similarities in costs and behavior in Canada and abroad. In either case the domestic industry has developed independently of the tariff and its structure would be placed under no pressure by the tariff's removal (cement, liquid and powdered detergents). Thus, the consequences of the Canadian tariff for the efficiency of manufacturing industry in Canada and prophecies as to the results of lowering the tariff can be put too strongly. Nevertheless, it is clear, on general theoretical grounds and also from an examination of the sample, that the Canadian tariff in most cases provides an opportunity to charge high prices that is accompanied by inadequate performance of the industry.

## THE EXPORTS OF AMERICAN-OWNED ENTERPRISES IN CANADA

By A. E. SAFARIAN

*University of Saskatchewan*

In recent years there has been considerable discussion in Canada concerning the consequences of the substantial ownership of some segments of Canadian industry by residents of the United States.<sup>1</sup> Much of this discussion has suffered from, among other things, the lack of comprehensive information on many significant aspects of the operations of the companies concerned. This paper deals with only a limited area of this topic by presenting some new data on the exports of such companies and by testing a few hypotheses concerning the results. Particular attention will be paid to the view frequently expressed in Canada that, except for a few primary resource products, American-owned firms are unlikely to export and particularly unlikely to export to the United States. The material may be of some interest for studies of the efficiency of Canadian industry in general and for more comprehensive studies of foreign investment.

The data in this paper are from a questionnaire, based on some preliminary interviews and covering several topics, which was mailed late in 1960 to about 1,500 Canadian firms or enterprises owned by non-residents in the manufacturing, mining, and petroleum industries. After some deletions, 280 usable questionnaires were available, including the 227 referring to firms owned by residents of the United States which form the basis of this paper. The response rate to the questionnaire was much better for the larger and medium-sized firms than for smaller ones. About 7 percent of the total number of firms owned by nonresidents in the industries noted and 40 percent of the assets of all such firms in these industries are covered by the study. The variable response rate to specific questions is referred to in the text where relevant. The data are subject to the usual qualifications regarding private

<sup>1</sup> The most comprehensive data on direct investments are in the census by the U.S. Dept. of Com., *U.S. Investments in Foreign Countries*, A Supplement to the Survey of Current Business (1960), and Dominion Bur. of Statis., *Canada's International Investment Position, 1926-1954* (1956). More limited information has since been published in various issues of the *Survey of Current Business* and *The Canadian Balance of International Payments, and International Investment Position*. For other analyses of the effects of foreign ownership, see Irving Brecher and S. S. Reisman, *Canada-United States Economic Relations*, Royal Commission on Canada's Economic Prospects (1957), especially Chap. 8, and John Lindeman and Donald Armstrong, *Policies and Practices of United States Subsidiaries in Canada*, The Canadian-American Committee (1960).



questionnaire surveys, including the fact that there was no way to control the pattern of responses. The distribution of firms in the study does correspond fairly closely with the overall population according to such simple aggregative tests as age distribution and degrees of ownership. In terms of numbers, however, the larger and medium-sized firms are overrepresented, as is manufacturing relative to primary mining and petroleum, and very few unincorporated branches are included. It should also be added that, in contrast with official definitions of direct investment companies, firms which have a substantial but unconcentrated ownership abroad have been excluded. The data on exports were requested for the year 1959 or for "an average of recent years" where 1959 was not typical. It is important to note that the export data consist entirely of frequency distributions of the percentage of the output of individual firms which is exported. In order to emphasize typical patterns, these percentages have not been weighted by the value of exports for the firms involved or in any other way, although distributions by the size of the assets of the firms are shown. It is also worth emphasizing that the great preponderance of firms represented here are engaged in manufacturing and/or assembly. In broad terms, about 65 percent of the firms indicated they were mainly engaged in producing fully processed or manufactured goods, 12 percent indicated they were mainly engaged in assembly or/and manufacturing, 4 percent were mainly extractive, 7 percent mainly semifabricated products, 3 percent semifabricated and manufactured products, and 8 percent largely or fully integrated production. In terms of the major groups of the Standard Industrial Classification, the industries most frequently represented are the electrical products industries (40 firms), machinery (37), chemicals and products (35), transportation equipment (22), metal fabricating industries (24), foods and beverages (15), and paper and allied industries (11), with a further 9 firms overlapping some of the industry groups just listed.

Before considering the percentages of output exported by the particular firms in this study, it is well to note the absolute values of production and of exports for all American-owned firms in Canada.<sup>2</sup> The detailed census data for 1957 collected by the United States Department of Commerce indicate that American-owned enterprises in Canadian manufacturing, mining, and petroleum industries had overall sales of U. S. \$10.7 billion. Fully 20 percent of these sales were for export; namely, \$1.4 billion to the United States and \$0.7 billion to other countries. Exports from Canada originating with these direct-investment companies accounted for about 50 percent of total Canadian exports of manufactures (including pulp and paper) in 1957, while for petroleum

<sup>2</sup> The absolute figures are from *U.S. Investments in Foreign Countries*, pp. 68, 110-111, and 114. Ownership and control ratios were derived from *The Canadian Balance of International Payments*, 1960, pp. 82-83. More recent data have not been published for some of the estimates used in this paragraph.

and other minerals and metals combined the proportion was over 85 percent. These proportions are higher than the proportion of capital invested in Canadian industry which was owned and controlled by residents of the United States at that time, as defined and measured by the Dominion Bureau of Statistics. At the end of 1957, residents of the United States owned 39 percent and controlled 43 percent of the capital of Canadian manufacturing industries, while American ownership and control of capital invested in petroleum and mining combined were 53 percent and 63 percent, respectively. These estimates suggest that, on an aggregative basis, the American-owned sector of the industries involved is more oriented to exports than are the other sectors. Such aggregative comparisons from different sources are open to a number of statistical objections, however. It is also well known that a few large firms in a few primary-resource oriented industries dominate the absolute values for exports; thus, \$1 billion of the total of \$1.4 billion exported to the United States in 1957 by American-owned firms consisted of metals and minerals and newsprint and pulp. The analysis of typical export patterns requires finer breakdowns.

### *I. Export Patterns*

The data presented here give some information on three points; namely, the proportion of companies exporting and the proportions of output exported, the geographical distribution of exports, and the portion of exports which represented trade with affiliated companies outside Canada.

About half of the 227 American-owned firms in the present study did not export in 1959. Exports accounted for 5 percent or less of sales in about one-quarter of the firms, while one-fifth of the firms indicated exports in excess of 5 percent of total sales. There is a marked increase in the ratio of exports to total sales as the size of the firm increases; the proportion of firms with no exports falls from 80 percent for those with assets under \$1 million to 54 percent for those with assets from \$1 to \$5 million, 13 percent for those with assets from \$5 to \$25 million and zero for those with assets over \$25 million. (Each of the four firms with assets over \$25 million, for which precise data were not available, had some exports in 1959.) It has already been noted that the method used diminishes the role of the few very large exporters of primary-resource products in order to show typical export patterns. It may be of interest, however, to note also the distribution in Table 1 of those 14 percent of the firms which indicated their type of business was mainly extractive or the production of semifabricated goods.<sup>3</sup> These

<sup>3</sup> The other types of business, and the proportions involved, were indicated above. The types shown here should not be confused with the distinction sometimes made between primary and secondary manufacturing. It should also be added that some awkward matters of definition are raised by all such classifications, particularly where multiproduct firms and various degrees of vertical integration exist.

TABLE 1

PROPORTION OF SALES IN CANADA, RESPONDENT AMERICAN-OWNED COMPANIES, 1959  
CLASSIFIED BY SIZE OF ASSETS  
(Number of Companies)

PERCENTAGE OF SALES IN CANADA	SIZE OF ASSETS IN MILLIONS OF DOLLARS				
	Under 1	1-4.9	5-24.9	25 and Over	All Companies
100.....	63	39	4	0	110
95-99.....	10	21	15	11	62
70-94.....	3	9	8	8	31
Below 70.....	4	2	1	7	15
Not available or no response to question.....	0	1	3	4	9
Totals.....	80	72	31	30	227

NOTE: The data for All Companies include 14 companies not classified by size of assets.

firms comprised 8 of the 15 exporting over 30 percent of output and dominated the firms in this group with assets over \$25 million. They comprised only one-sixth of those exporting from 5 to 30 percent of output, however, and did not predominate in any size group. They comprised 10 percent of each of the other groups shown and were concentrated among the two smaller size groups in these cases.

It is sometimes suggested that when American-owned companies do export (again excepting a few large exports closely related to primary resources) they are more likely to export to overseas markets than to

TABLE 2

DESTINATION OF EXPORTS, RESPONDENT AMERICAN-OWNED COMPANIES WITH EXPORTS, 1959  
(Percentages by Each Destination)

PERCENT- AGE OF EXPORTS TO DESTINATION INDICATED	DISTRIBUTION OF RESPONSES BY EACH DESTINATION							
	Area				Affiliation			
	U.S.	U.K.	Other Sterling Area	Other	Parent	Other Foreign Affiliates of Parent	Own Subsidi- ary Abroad	Other
100.....	24	2	13	7	13	4	0	33
70-99.....	20	2	3	10	11	1	3	21
30-69.....	11	5	3	12	10	6	0	10
1-29.....	14	25	24	18	18	10	0	11
0.....	31	66	56	54	48	61	9	26
Not relevant						18	88	
	100	100	100	100	100	100	100	100

NOTE: The data should be read by columns, which total 100 percent except for rounding. The cases shown as not relevant refer to companies with exports but where the parent companies had no other foreign affiliate or the Canadian company did not have a subsidiary outside Canada. About 90 companies are involved in this table (after excluding those with exports in Table 1 which could not give destinations) of which 17 indicated their type of business was extractive or semifabricated products.

the United States, and particularly to the United Kingdom and other sterling markets where tariff preferences exist. The data in Table 2 suggest that companies which do export are more likely to export to the United States market. Thus, 34 percent and 44 percent of those companies with exports had some exports to the United Kingdom and to other sterling area countries, respectively, while 69 percent of those with exports did some exporting to the United States. The most likely export destination was to nonaffiliated entities, whether in the United States or elsewhere; 74 percent of the firms with exports reached such destinations. Companies with exports very frequently found markets with their affiliates abroad, however. About half of the companies with exports reported some exports to the parent company in the United States; indeed, fully one-third of those with exports sold between 30 and 100 percent of their exports to the parent company. A smaller portion of the firms with exports also sold to affiliates of the parent in third countries.

## II. *Ownership and Exports*

One of the most persistent criticisms of foreign ownership of Canadian industry is that the firms involved tend to export less than they might simply because they are owned in part or whole by nonresidents. Since some of these firms export a very substantial volume of primary or semifabricated products, as already noted, the argument is usually put in terms of exports of manufactures not directly related to a significant natural resource base.

This idea has found growing expression in Canada over the past few years, particularly as a result of views expressed to and by the Royal Commission on Canada's Economic Prospects.<sup>4</sup> It was frequently echoed in the speeches of Ministers of the past government and is very often mentioned by those of the present government. It has been quite explicitly advanced as part of the reason for current legislation which, briefly, will raise withholding taxes on dividends paid abroad by firms which have less than a stated degree of Canadian ownership and reduces such taxes for other firms. The expectation from such legislation is that some firms now largely or wholly owned by nonresidents will avoid the tax increases by meeting the requirements of the act, and that the effect of a significant minority shareholding and of a minimum number of resident directors will be to encourage the use and development of domestic production and other facilities, including export capacity.<sup>5</sup>

<sup>4</sup> See the *Final Report*, Royal Commission on Canada's Economic Prospects (1957), Chap. 18. For an earlier statement, see the Report of the Commissioner, Combines Investigation Act, *Canada and International Cartels* (1945), part II.

<sup>5</sup> The precise nature of the tax changes and the reasoning underlying them can be found in Canada, *House of Commons, Debates*, 26 Parliament, 1st Sess. (1963), June 13 (especially p. 1001), July 8, and Oct. 16 (especially pp. 3638-39). In the budget presented on March 16,

Much could be said about these specific provisions, but only the general position concerns us here. Just why ownership as such should limit exports has never been clearly specified. There appear to be two loosely articulated positions, however. One is that the existence of facilities in other countries, owned by the same international firm, must inhibit exports because the sales may be at the expense of a related firm. Assuming rational economic behavior on the part of the international firm, any such limitation on competitive exports would apply in the short run only. Other special assumptions are also involved, as noted in the analysis of this point in the study by Brecher and Reisman. The other position which is sometimes voiced is that export limitation reflects noneconomic factors, such as international market allocation within the international firm for the sake of administrative convenience, various private and public pressures and preferences favoring production in the parent's country, and ignorance of foreign costs on the part of officials of the parent company. It is extremely difficult to identify and to assess such noneconomic factors. One can only suggest that their persistence would imply a disregard for profit maximization and that there is no logical reason why a Canadian-owned firm with an international interest would be immune from similar factors.

The present study did not attempt a direct measure of the prevalence of private export restrictions within the international firm but did include a question regarding the effects of affiliation with companies abroad on the export volume of the subsidiary. The question is both general and subjective, but it elicited a response from almost two-thirds of those returning the questionnaire. About half of those who answered the question indicated either that affiliation had little or no effect on exports or that restrictions as such were absent; 20 percent indicated the question was not applicable for various reasons, such as recent establishment in Canada and the unsuitability of the product for export; 20 percent indicated the effect on exports was favorable, often highly favorable, for such reasons as guaranteed sales and other contacts abroad; and 10 percent reported private export restrictions or that affiliation reduced exports, at least with regard to some portion of the firm's output. No doubt such a broad question needs to be supplemented with more intensive studies. For example, some aspects of pricing practices through international export corporations and of the allocation of patents deserve more study to determine whether they are actually or potentially restrictive and whether they can be renegotiated where they do impede exports which are otherwise competitive. Such studies, to be

---

1964, it was announced that the increase in withholding tax will not go into effect, although the decrease will still apply for firms which have a stated degree of Canadian ownership.

realistic, must appreciate that private restrictions may have less to do with ownership than with the abuse of legal monopolies, as in the case of international market allocation through patents. It should also be noted that the effect of such restrictions often is to protect the Canadian firm at least temporarily from the competition of foreign affiliates; some discussion on this point in Canada assumes that the protection runs only in the reverse direction. Whatever more intensive studies may reveal, one must note that the overwhelming impression of the present study is not one of restriction by the parent company but rather of inability to compete on the one hand and of considerable advantages in exporting on the other. Indeed, it is worth emphasizing that the subsidiary company frequently has certain advantages in this respect. Its unit costs tend to be reduced by access (most often at less than a market price) to the research, development, and other facilities of the affiliate; and should it be or become competitive, it often is able to acquire the advantage of an existing distribution system through its affiliate.

Those who suggest there is a relationship between ownership and exports have gone so far as to suggest there is a relationship between the degree of foreign ownership and exports. As noted earlier, the basis for the discriminatory changes in withholding taxes is, in part, the expectation that a significant minority resident shareholding (and resident directors) will tend to raise the proportion of output which is exported. I have compared the degree of resident ownership of the companies in this study with the proportion of output which is exported using simple correlation techniques.<sup>6</sup> There is no support in this data for the view that a minority resident ownership tends to be associated with an increased share of exports in output, whether for all companies or for those with assets over \$25 million.

The most complete test of the view that ownership and exports are related would be a comparison of the proportion of output exported by resident-owned and nonresident-owned firms in which all other relevant variables were held constant. An adequate test of this kind would require a much larger number of firms than it has been possible to survey in the present study, and even then there would be complications because of the great variation in nonresident ownership of particular industries and in the export-determining characteristics of firms. Nevertheless, a limited test can be made by means of a questionnaire which was sent to resident-owned firms in those industries where both resident and nonresident ownership is significant. This test has been limited to

<sup>6</sup> All 280 companies were used for this purpose including both those owned by residents of the United States and those owned by other nonresidents. Multiple correlation techniques do not appear feasible here and in the correlation with unit costs noted below, partly because the data tend to cluster at certain points (100 percent foreign ownership, 100 percent of sales in domestic market) and, for the remainder, the combined observations are reduced by differential response rates to different questions.

firms with assets over \$1 million in each case and to the major commodity group classification of the Standard Industrial Classification. Because of uneven numbers of firms and of responses by industry in the two ownership categories, comparisons have been limited to industries where at least four firms were available in each ownership category. It need hardly be emphasized that the economic and statistical constraints mentioned here are exceedingly simple; yet even these have reduced by half the number of industry groups for which comparisons were possible and to 70 and 113, respectively, the number of resident-owned and nonresident-owned firms in the comparison. For each of the 10 remaining broad industry groups the medians for the percentages of output exported by individual firms were compared. In 1 industry group the medians were identical, in 4 the medians were higher for the nonresident-owned firms and in 5 they were higher for the resident-owned firms. In only 1 industry did the difference in the median export percentages for the 2 groups of firms exceed 7 percent, and in 7 of the 10 industries it was 5 percent or less. Admittedly, such comparisons need to be extended and refined.<sup>7</sup> As they stand, they suggest that any problems which inhibit Canadian exports are generally common to both sets of firms, and that in the typical case ownership as such is not a significant deterrent to the exports of nonresident-owned firms. We must turn to more conventional and more convincing explanations of export patterns, if only to note the outlines of some of the problems.

### III. *Relative Unit Costs and Exports*

For an economic analysis of export patterns of American-owned companies it would be most interesting to have, among other things, data on relative rates of return of various geographic locations. Data on unit costs of production were more readily available, though there were problems of collection here, too. The companies were asked to compare their unit costs of production for their major comparable products with those of the parent company abroad, at the current rate of exchange and at normal volume of operations. About two-thirds of the companies returning the questionnaire answered this question; 55 percent of those answering the question said unit costs were typically higher than those of the affiliate, 20 percent that they were about the same, 12 percent that they were typically lower, and about 12 percent that relative unit costs varied on major products or that such products

<sup>7</sup> A much narrower definition of industry would be desirable for this test, given more firms and some disaggregation to the level of the establishment. Where a number of firms showed substantial variations from the median for the industry, a closer check of the data usually revealed that the variations were due to the broad nature of the industry groups used and that they occurred for both sets of firms. In the paper and allied industries, for example, both the resident-owned and nonresident-owned firms producing newsprint and pulp showed very high export ratios while both types of firms showed low export ratios for other paper products. Wide variations from industry medians by individual firms also appear to follow product differences rather than the nature of ownership.

were not comparable.<sup>8</sup> Analysis of the distribution of responses by size of assets of the subsidiary is conditioned by the few responses in some categories. It may be noted, however, that, disregarding those for which such comparisons were not reported, the proportion with unit costs typically higher than those of the affiliate is about 60 percent for each size group under \$25 million but just over 20 percent for those with assets over \$25 million. The larger companies also have a correspondingly greater proportion of cases where the relative unit cost varies or the products are not comparable.

Companies with unit costs which were typically higher or lower than those of the affiliate were asked to indicate the approximate average difference. Not unexpectedly, the response rate fell further here. Of the 45 companies with higher costs which answered this question, 20 had unit costs ranging up to 10 percent in excess of those of the affiliate and a further 20 had unit costs from 11 to 20 percent higher than those of the affiliate. The 7 companies indicating the extent to which their costs were lower were all within 15 percent of the affiliate's costs. The range of cost variation is not substantial, judging by these data, though one would feel more secure in such a conclusion from this particular question if it had been answered by more of the companies returning the questionnaire. The only other recent general study on this, covering fewer companies but much more intensively, arrived at conclusions similar to those just noted.<sup>9</sup>

TABLE 3

UNIT COSTS OF PRODUCTION RELATIVE TO AFFILIATE, RESPONDENT AMERICAN-OWNED COMPANIES IN CANADA, 1959  
(Number of Companies)

UNIT COST RELATIVE TO AFFILIATE	SIZE OF ASSETS IN MILLIONS OF DOLLARS				
	Under 1	1-4.9	5-24.9	25 and Over	All Companies
Higher.....	32	27	12	5	80
About the same.....	11	11	3	4	29
Lower.....	9	5	2	2	18
Varies and not applicable.....	1	2	4	11	18
No response and not available..	27	27	10	8	82
Total.....	80	72	31	30	227

NOTE: The data for All Companies include 14 companies which are not classified by size of assets.

<sup>8</sup> Since the unit cost comparisons apply only to comparable products, the extent to which they represent the overall output of the subsidiaries will depend in large part on the degree of similarity of the products of the two sets of companies. The comparisons appear to be fairly representative. Replies to other parts of the questionnaire suggest that, in general, the products are identical with those of the parent or only marginally different, except for a small minority of companies. Furthermore, where the products are comparable, the great majority of the companies are producing at least a majority of the parent's range of products.

<sup>9</sup> See Theodore R. Gates and Fabian Linden, *Costs and Competition: American Experience Abroad* (Nat. Ind. Conf. Bd., 1961), especially Chaps. 2 and 7.



It is interesting to note the relationship between unit costs relative to the American affiliate and the proportion of the subsidiary's output which is exported to all markets. Sixty percent of those firms which reported unit costs higher than those of the affiliate also reported no exports; the corresponding proportion was 40 percent for firms with unit costs the same as, lower than, or variable compared with the affiliate. About 15 percent of the firms with unit costs higher than those of the affiliate were still able to export in excess of 5 percent of their output; the corresponding percentage for the other group of firms was close to 30 percent.<sup>10</sup> Obviously one must not claim too much for such comparisons; the products involved face quite different tariff and other obstacles to exports, distribution costs are not included, and so on. At the very least, the data are not inconsistent with the view that relative competitive ability (as measured here) plays a role in the export patterns of subsidiary companies.

The reasons for cost differences, as reflected in this study, can only be summarized here. Eighty-four companies gave reasons for the cost differences, about a third of them giving more than one reason, yielding 117 observations. Lower wage rates in Canada made up 12 of the 19 observations for companies with lower unit costs. The 98 responses for companies with higher unit costs were concentrated in a few categories. One group of 55 responses referred explicitly to a relatively lower volume of operations, shorter production runs, and similar answers. A further 15 observations referred explicitly to import duties as the basic reason for higher unit costs. Still another 14 referred to the higher cost of materials in Canada, several of which clearly involved imported materials. It might be noted that the other general study on unit costs, referred to earlier, places major emphasis on relatively higher costs of materials and components in Canada.

The meaning to be attached to the responses emphasizing the volume of operations depends, of course, on broader considerations with regard to the economic environment and public policy.<sup>11</sup> Indeed, the findings of this study generally are within a given context of industrial structure, tariff and exchange rate policies, and other variables. Consideration of the effects of changes in these is beyond the scope of the present paper, but perhaps enough has been said to indicate some areas worth exploring for those interested in the exports of subsidiary companies and the broader question of the efficiency of Canadian industry.

<sup>10</sup> These contrasts are greater still if one compares firms with unit costs higher than those of the affiliate only with those having unit costs lower than those of the affiliate, but the latter are so few in number that the results are probably not significant.

<sup>11</sup> The most frequent explanation of the causes and effects of low volume is treated extensively in D. H. Fullerton and H. A. Hampson, *Canadian Secondary Manufacturing Industry*, Royal Commission on Canada's Economic Prospects (1957). For an alternative explanation, see the paper by H. C. Eastman presented to this session.

# THE U.S. TARIFF AND THE EFFICIENCY OF THE U.S. ECONOMY\*

By ROBERT M. STERN  
*University of Michigan*

The purpose of this paper is to evaluate for the United States the economic significance of its tariffs and import quotas. The procedure followed is to estimate for selected postwar years the increases in United States imports from all countries and particularly from Canada on the assumption of suspension of existing tariffs and quotas. The estimated increases in imports are then used as a basis for calculating the gains in economic welfare that such suspension might provide. The relatively small magnitudes of the estimates both of the import increases and welfare gains in relation to national product and income suggest that the American economy could probably adjust without serious difficulty to the complete liberalization of imports and of trade generally.

## I

In order to provide perspective on what is to follow, it is worth noting that even though United States merchandise imports in 1962 amounted to \$16.2 billion, making this nation the world's single largest importer, these imports were only 2.9 percent of gross national product. Moreover, \$6.2 billion (38.3 percent) of these imports were free of duty. While the \$10.0 billion of dutiable imports yielded in excess of \$1 billion in customs revenues, it should be noted that there was substantial variation in the tariff rates applicable to imports of individual commodities and commodity classes.

While details on average tariff rates for 1962 dutiable imports were not readily available, for 1956 these rates were approximately 6 percent for nonagricultural crude materials and semimanufactures, 8 percent for crude foodstuffs, 12 percent for manufactured foodstuffs, 18 percent for agricultural crude materials,<sup>1</sup> and 20 percent for finished manufactures [13, pp. 213-14].<sup>2</sup> The levels of United States duties can thus

\* I would like to acknowledge the many helpful comments received on an earlier version of this paper from colleagues and from members of the International Economics Seminar at Michigan. Diane Welch is to be thanked for her aid in computations. The research underlying this paper was made possible with assistance received from the University of Michigan Ford Foundation Faculty Research Fund.

<sup>1</sup> The relatively high rate on agricultural crude materials was attributable chiefly to wool. Rates on other materials were relatively much lower.

<sup>2</sup> Newsprint and burlap are regarded here as semimanufactures rather than finished manu-

be seen to increase more or less with the degree of processing or fabrication to which the imports have been subjected. Within the finished manufactures category, there is also a pronounced tendency for rates to vary directly with the importance of labor costs relative to total costs, so that, for example, rates on textiles and apparel are significantly greater than rates on machinery and vehicles.<sup>3</sup>

In order to determine the increase in imports of a particular commodity and of a wide variety of commodities in the event of tariff and quota suspension, we evidently need information on demand and supply conditions at home and abroad and on the relevant substitutions and complementarities involved. Since, however, it would take an enormous research effort to obtain the requisite data, simplifications in terms of assumptions and method are evidently necessary.

Howard S. Piquet's estimates [7] of the increases in 1951 imports as a consequence of a unilateral, temporary suspension of tariffs and quotas provide us with a convenient starting point. His estimates were based upon "expert" opinion concerning the increases in imports that might be realized in a three- to five-year period during which market conditions and prices prevailing in 1951 would be continued. In order to provide a range of estimates, the increases were expressed as minimum and maximum percentages of actual 1951 values.<sup>4</sup>

Since Piquet's sample of dutiable imports included 226 commodities or commodity groups, it was thought desirable for present purposes to allocate the individual items according to the Census of Manufactures

factures. The importance of the various import classes in 1956 can be judged from the following percentage distribution [13, p. 249]:

Economic Class	Dutiable Imports	Free Imports	Total Imports
Crude materials.....	21.1%	28.2%	24.6%
Crude foodstuffs and animals.....	3.5	29.2	16.3
Manufactured foodstuffs.....	18.1	.5	9.3
Semimanufactures.....	26.2	34.2	30.2
Finished manufactures.....	31.1	7.9	19.6
Total.....	100.0	100.0	100.0
Value (millions of dollars).....	\$6,271	\$6,220	\$12,491

<sup>3</sup> For further details which refer to 1954, see [14, especially pp. 50-66]. Note that the average tariff rates cited should not be taken literally as indicating the restrictiveness of the tariff, because of the well-known downward bias created by weighting the rates with the actual value of imports and the lack of information concerning the degree of protection which the tariffs in fact provide.

<sup>4</sup> In constructing his estimates, Piquet made allowances for substitutions, complementarities, and bottlenecks. His sample covered 80 percent of dutiable imports. The 20 percent which were not covered were assumed to increase in the same proportion as the 80 percent. For further discussion of these and other assumptions, see [7, pp. 11-20].

TABLE 1  
ESTIMATED EFFECTS OF SUSPENSION OF TARIFFS AND QUOTAS ON SELECTED DUTIABLE IMPORTS INTO THE UNITED STATES, 1951 AND 1960  
(Millions of Dollars)

SIC	COMMODITY GROUP	(1) (2)		(3)	(4) (5)		(6) (7)		(8)
		VALUE OF IMPORTS			ESTIMATED INCREASE IN IMPORTS IN EVENT OF		TARIFFS BASED UPON		
		1951	1960		TARIFF SUSPENSION ONLY Based Upon		QUOTAS BASED UPON		
				Piquet d% Method 1951	Import Elasticity Method 1951	Piquet d% Method 1960	Piquet d% Method 1951	Import Elasticity Method 1951	Piquet d% Method 1960
01 09	Selected dutiable imports (226 commodities) Agricultural commodities..... Fisheries commodities.....	795 6	211 4	611	92	96	681	162	274
10 14	Metal mining products..... Nonmetallic minerals, nec.....	801	215	611	92	96	681	162	274
		93 25	292 33	17 10	4 5	22 33	17 10	4 5	22 33
		118	325	27	9	55	27	9	55
20 21 22 23 24 26	Food and kindred products..... Tobacco manufactures..... Textile mill products..... Apparel and related products..... Lumber and wood products..... Paper and allied products.....	862 82 285 38 240 21	1,290 103 363 78 341 31	220 37 116 23 8 11	211 51 408 125 59 38	293 39 195 48 15 22	983 37 116 23 8 11	974 51 408 125 59 38	1,262 39 212 48 15 22

(Continued on next page)

NOTES: Columns 1, 3, and 6 based upon data in Piquet [7] classified according to the commodity codes in [11]. Column 2 based upon the commodities covered for 1951 as recorded in [10]. Column 4 estimates based upon computed price elasticities of import demand; see text and footnotes 6 to 8. Column 5 estimates based upon application of maximum percentage increases for 1951 as listed in [7] to 1960 dutiable imports. Column 7 is column 4 plus the estimated increase in 1951 imports due to quota suspension (column 6-3). Column 8 is column 5 plus the estimated increases in 1960 imports due to quota suspension; see footnote 11.

TABLE 1—(Continued)

SIC	COMMODITY GROUP		(1)		(2)		(3)		(4)		(5)		(6)		(7)		(8)	
			VALUE OF IMPORTS		1960		Tariff Suspension only Based Upon		Suspension of Tariffs and Quotas Based Upon		Piquet d/% Method 1960		Piquet d/% Method 1951		Import Elasticity Method 1951		Piquet d/% Method 1960	
	1951						Piquet d/% Method 1951	Import Elasticity Method 1951	Piquet d/% Method 1951	Piquet d/% Method 1960			Piquet d/% Method 1951	Import Elasticity Method 1951			Piquet d/% Method 1960	
27	5				14		5	10		14			5		10		14	
28	132				91		52	97		26			52		97		26	
29	534				1,434		48	267		153			48		267		143	
31	42				105		110	129		279			110		129		153	
32	37				96		49	53		89			49		53		279	
33	365				543		6	100		18			6		100		126	
34	3				13		1	19		1			1		19		18	
35	1				1		15	35		35			15		35		1	
36	15				35		21	85		76			21		85		35	
37	39				625		42	68		91			42		68		76	
38	74				98		20	154					20		154		91	
39	100				181													
	2,875				5,442		784	1,910		1,707			1,547		2,673		2,873	
	3,794				5,982		1,422	2,011		1,858			2,255		2,844		3,202	
	1,030				2,888		386	546		897			386		546		897	
Total.....																		
Other dutiable imports.....																		
Total dutiable imports.....	4,824				8,870		1,808	2,557		2,755			2,641		3,390		4,099	
Total free imports.....	5,993				5,780													
Total dutiable and free imports.....	10,817				14,650													
Estimated cost of protection.....							113	183		159			238		308		349	

Standard Industrial Classification (SIC).<sup>5</sup> The resulting classification of the \$3.8 billion of dutiable imports is shown in column 1 of Table 1. The estimated maximum increases of \$1.8 and \$2.6 billion in the event of tariff suspension and of tariff and quota suspension together, which Piquet obtained by application of his percentage method, are shown in columns 3 and 6.

In view of his assumption of a temporary suspension of tariffs and quotas, Piquet's estimates are perhaps on the low side. This is because too little allowance is made for the stimulative effect which a permanent tariff suspension might have on the foreign supply of exports. In light of this factor and also because of the explicit judgmental character of his work, it was felt that it would be of interest to construct on the basis of somewhat different assumptions alternative estimates of increases in imports in the event of tariff suspension. It was attempted, accordingly, to approximate the increases in dutiable imports using price elasticities of import demand,  $e_m$ , computed according to the formula:  $e_m = (Q_d/Q_m)e_d - (Q_d/Q_m)e_s$ , where  $Q_d$  refers to total demand, including imports;  $Q_s$  to domestic supply;  $Q_m$  to dutiable imports; and  $e_d$  and  $e_s$ , respectively, to the price elasticity of total demand and domestic supply.<sup>6</sup>

In the absence of direct estimates of  $e_d$  and  $e_s$ , representative elasticities were assumed to be  $-.25$  and zero for crude materials, foodstuffs, and animals,  $-.40$  and  $.20$  for semimanufactures,  $-.50$  and  $.25$  for nondurable finished manufactures, and  $-1.00$  and  $.50$  for durable finished manufactures.<sup>7</sup> In dealing with cases where imports were a small fraction of domestic production and consumption and where tariffs might possibly be unusually restrictive of imports, an import demand elasticity of  $-25.0$  was assigned in order to provide a generous margin for increases in imports.

Then, on the assumption that for each commodity, the tariff,  $t$ , was reflected fully in the domestic price,<sup>8</sup> the proportion of the price represented by the tariff,  $t/1+t$ , was multiplied by  $e_m$  to obtain the estimated percentage increase in imports. Each percentage increase was multiplied

<sup>5</sup> The 1951 imports were grouped according to the list of output commodity classes based on the 1957 SIC and constituent Schedule A import commodity codes, as shown in [11, pp. 59-82].

<sup>6</sup> For a derivation of this formula, see [6, p. 127].

<sup>7</sup> The demand elasticities were chosen in the light of the relatively low elasticity estimates which statistical studies have in general produced. See [5, esp. p. 335, footnote 19]. The supply elasticities were varied according to judgment concerning the comparative ease or difficulty with which the different economic sectors could expand or contract output. It may be noted that a halving or doubling of the assumed elasticities would affect the value of the estimated increases in approximately the same proportions.

<sup>8</sup> This is equivalent to assuming that imports are obtained at constant costs, and further implies that increases in quantity and value of imports can be taken to be equal. If, on the other hand, costs are assumed to be increasing, the extent to which the domestic price reflected the tariff would depend upon the magnitudes of the home and foreign elasticities of demand and supply.

by the corresponding value of 1951 imports, and the results were summed over all commodities according to SIC groupings.

Details of the total estimated increase of \$2.6 billion in 1951 imports obtained by application of the foregoing method are shown in column 4 of Table 1. Retaining Piquet's estimates of increased imports due to quota suspension, the alternative calculations yielded a \$3.4 billion increase in 1951 imports (column 7). It should be noted that my total estimate in column 4 is not only larger than Piquet's in column 3, but that the details also differ because of the variations in the computed import demand elasticities. The chief differences lie in agricultural commodities, where computed elasticities were relatively low, and in textiles, apparel, petroleum, and miscellaneous manufactures, where the computed elasticities were relatively high. Too great reliance upon the details should be avoided, however, in view of the simplifications which have been employed.

For purposes of comparison and to utilize data of more recent vintage, estimates using Piquet's percentage method were made for comparable 1960 imports.<sup>9</sup> These estimates are shown in columns 5 and 8 of Table 1. The 1960 imports of the commodities included in Piquet's sample were \$6.0 billion out of a total of \$8.9 billion of dutiable imports (column 2). Applying Piquet's estimated 1951 maximum percentage increases to 1960 revealed a possible rise of \$2.8 billion in 1960 imports in the event of tariff suspension. If both tariffs and quotas were suspended, the estimated rise in 1960 imports becomes \$4.1 billion. Unfortunately, because detailed production data were not readily available, import demand elasticities could not be computed for 1960. But if the ratio of the column 3 and 4 total increases in 1951 imports due to tariff suspension is applied to the 1960 total in column 5 the increase in 1960 imports is an estimated \$3.9 billion.<sup>10</sup> Taking quota suspension also into account gives an estimate of a \$5.2 billion increase in 1960 imports.<sup>11</sup>

Judged in relation to dutiable imports, the estimated increases in the event of tariff and quota suspension were 55 and 70 percent of 1951 imports and 46 and 59 percent of 1960 imports. Estimated increases in relation to total imports were 24 and 31 percent for 1951, and 28 and 36 percent for 1960. While these percentages and the associated dollar

<sup>9</sup> In order to obtain comparable data for 1960 imports, an attempt was made to match the 1951 Schedule A import commodity codes underlying Piquet's data in [12] with the 1960 codes recorded in [10].

<sup>10</sup> This estimate is close to one of \$4 billion applied to 1956 imports by Piquet in [13, p. 254], on the basis of extrapolating an unofficial Treasury Department guess for an earlier year.

<sup>11</sup> The assumed effects of 1960 quota removals include estimates for rye, almonds, lead, zinc, cotton yarn, countable cotton cloth, petroleum, and wool in addition to the commodities covered by quotas in 1951. Although wool imports were actually not subject to quota in 1960, imports would likely increase with a suspension of deficiency payments to wool growers. The point just made should be a reminder that removal of quotas implies discontinuance of the associated domestic programs, many of which were adopted for reasons other than protection alone.

magnitudes are hardly insubstantial, it should be noted that even a \$5.2 billion increase in 1960 imports would be scarcely more than 1 percent of gross national product.

On the basis of the foregoing estimates, we can try to assess for the American economy the gains in economic welfare from the suspension of tariffs and quotas. We have here to draw upon the Marshallian consumer surplus concept and attempt to estimate what Corden [2] and Johnson [5] have termed the "production" and "consumption" costs of protection.<sup>12</sup> Using, then, the same assumption as above that the tariff was reflected fully in the domestic price of each commodity, the estimated increase in the value of imports due to tariff suspension was multiplied by one-half of the corresponding ad valorem tariff rate and summed over all commodities. Comparable estimates for quota suspension were made, based upon differences between domestic and world prices of the relevant commodities. The sum totals of these computations are shown at the foot of columns 3 to 8 in Table 1.

The possible increases in economic welfare, as measured by the foregoing computations of the cost of protection, were \$113 and \$183 million in the event of tariff suspension and \$238 and \$308 million in the event of tariff and quota suspension for 1951. The estimates for 1960 were \$159 and \$349 million, which, if increased by the ratio of the 1951 figures in columns 3 and 4, would be \$258 and \$448 million. These estimates are striking in view of their relatively small magnitudes. The largest of them in 1951 and 1960 equaled only .11 percent of national income in the respective years. It thus appears that the welfare increase consequent upon tariff and quota removal would doubtlessly remain a comparatively small proportion of national income even if estimated increases in imports were significantly larger than those indicated.<sup>13</sup>

The point, therefore, of the foregoing exercise is to demonstrate the

<sup>12</sup> Strictly speaking, we are dealing here with a static, partial equilibrium concept, which assumes constant-utility demand curves, no cross-effects with respect to other goods, constant terms of trade, and the absence of employment and balance-of-payments considerations. Furthermore, in measuring the production and consumption costs of protection, no account is taken of the other portions of the increase in consumer surplus which consist of the redistribution of income from producers to consumers and of the loss of tariff revenues.

<sup>13</sup> Denison [3, pp. 190-91] has estimated roughly, on the basis of Piquet's data, a cost of protection equal to .17 percent of national income in 1951. If account were taken of the United States restrictions covering tourism, shipping, and other services, the "Buy American" Act, "voluntary" quotas, tied aid, and federal purchase requirements, he estimated that the gain from the elimination of trade barriers might well be three or four times the figure stated. But in any event, as he argues, trade barriers have been insignificant in their impact upon the past rate of United States economic growth, and it is unlikely, therefore, that trade liberalization would be of much consequence for this nation's future growth.

A more complete analysis would of necessity have to make allowance explicitly for the relevant substitutions and complementarities. Moreover, consideration would have to be given to the effects of increasing costs of imports, export expansion, changes in the terms of trade, and the possible adjustments required for maintenance of full employment and balance-of-payments equilibrium. While this latter group of factors would present serious measurement problems, their effect, as Corden [2, pp. 36-47] especially has pointed out, might be to yield estimates of the cost of protection even smaller than those indicated.



relatively small cost to the American economy of the protection which tariffs and quotas provide and to suggest that import liberalization could be achieved without serious adjustments in the economy as a whole.<sup>14</sup> If account is taken, moreover, of the likelihood of a multilateral rather than a unilateral suspension of tariffs and quotas, the resultant increase in this nation's total production and exports would certainly enable resources to be reallocated all the more easily.

## II

Before considering the question of American import liberalization with respect to Canada,<sup>15</sup> we may note that in 1961 imports into the United States from Canada totaled \$3.1 billion as compared with total United States imports from all countries of \$14.7 billion. Of the \$3.1 billion of Canadian exports to the United States, 61 percent, or \$1.9 billion, entered duty free. Of the remaining \$1.2 billion, \$754 million were dutiable at 5 percent or less, \$385 million at from 6 to 19 percent, and \$89 million at 20 percent and over [8, Chart 3]. Canada's exports to the United States and to all countries for 1961, classified on the SIC basis, are shown in columns 1 to 4 of Table 2. The predominance of the United States, which took 54 percent of Canada's 1961 exports, and the concentration on exports of crude and semiprocessed agricultural, forestry, and mineral products are clearly evident in the table.

Some indication of the possible impact of trade liberalization on American imports from Canada is available from other work of Howard S. Piquet [8]. Using his percentage method, he estimated on the basis of 1961 data an increase of \$262 million or more in imports from Canada on the assumption that full use was made of Presidential powers to liberalize imports under the Trade Expansion Act of 1962.<sup>16</sup> In contrast, assuming complete and permanent removal of United States tariffs and quotas with respect to Canadian goods and applying a combination of the elasticity and percentage methods, I obtained an estimated increase

<sup>14</sup> Using Salant and Vaccara's 1953 estimate [9, p. 215] of a median gross decrease in employment of 115 employees per million dollars increase in imports, it may be noted that approximately 304 or 390 thousand employees would require re-employment in connection with the estimated increases in 1951 imports and 471 or 603 thousand with the 1960 increases. This amount of employment displacement from imports does not appear unusually large in relation to the 2.2 and 2.0 million decline in employment in the 1953-54 and 1957-58 recessions. The estimated displacement would be smaller if account were taken of the employment effects of export expansion due to the increase in foreign incomes.

<sup>15</sup> This portion of the paper has purposefully been made brief in anticipation of Ronald J. Wonnacott's forthcoming study of the estimated impact of free Canadian-American trade upon the amount and location of production in the two countries.

<sup>16</sup> These powers include the elimination of tariffs of 5.0 percent or less, the elimination of tariffs on goods of which the United States and the European Economic Community supplied 80 percent or more of free world exports in 1960, and the reduction of remaining tariffs by up to 50 percent. Although it was outside the scope of his study, it would have been interesting if Piquet had estimated the increases in Canada's imports from the United States under the 80 percent authority especially.

TABLE 2  
ESTIMATED EFFECTS OF SUSPENSION OF UNITED STATES TARIFFS AND QUOTAS ON SELECTED CANADIAN EXPORTS, 1961  
(Millions of U.S. Dollars)

SIC	COMMODITY GROUP	(1) (2) (3) (4)			(5) ESTIMATED INCREASE IN CANADIAN DUTIABLE EXPORTS TO U.S. IN EVENT OF U.S.	(6) Quota Suspension (% Method)	Tariff and Quota Suspension		
		VALUE OF 1961 CANADIAN EXPORTS TO							
		All Countries	Total	Free	Dutiable			Tariff Suspension (Elasticity Method)	
01 08	Agricultural commodities..... Forestry commodities.....	865 6	115 6	12 6	103	88	24 112		
		871	121	18	103	88	24 112		
10 12 13 14	Metal mining products..... Bituminous coal and lignite..... Crude petroleum and natural gas..... Nonmetallic minerals, nec.....	346 8 194 335	126 2 194 227	104 2 41 227	22 153	2 46	4 46 52		
		883	549	374	175	48	4 52		
20 21 22 24 26	Food and kindred products..... Tobacco manufactures..... Textile mill products..... Lumber and wood products..... Paper and allied products.....	286 27 8 369 1,158	152 * * 274 926	21 * * 906	131 * * 20	205 * 30 22	3 * * 30 22		

(Continued on next page)

\* Less than \$500,000.

Notes: Columns 1 to 4 based upon data in Piquet [8] classified according to the commodity codes in [11]. Column 5 estimates based upon computed price elasticities of import demand; see text and footnotes 6 to 8. Column 6 estimates based upon 1960 estimates for quotas applicable to Canada; see footnote 11. Column 7 is the sum of 5 and 6.

TABLE 2—(Continued)

SIC	COMMODITY GROUP	(1) (2) (3) (4)			(5) ESTIMATED INCREASE IN CANADIAN DUTIABLE EXPORTS TO U.S. IN EVENT OF U.S.	(6) Quota Suspension (% Method)	(7) Tariff Suspension and Quota Suspension
		VALUE OF 1961 CANADIAN EXPORTS TO					
		All Countries	United States				
		Total	Free	Dutiable			
28	Chemicals and allied products.....	213	94	70	24	50	50
31	Leather and leather products.....	7	2		2	2	2
33	Primary metal products.....	849	363	33	330	87	93
34	Fabricated metal products.....	11	6		6	17	17
35	Machinery, excluding electrical.....	79	48	35	13	27	27
36	Electrical machinery.....	20	16		16	50	50
37	Transportation equipment.....	146	86		86	238	238
38	Instruments and related products.....	11	6		6	14	14
		3,184	1,973	1,065	908	742	751
	Electricity.....	16	16	16			
	Total selected Canadian exports.....	4,954	2,659	1,473	1,186	878	915
	Other exports.....	731	411	382	29	21†	21†
	Total Canadian exports.....	5,685	3,070	1,855	1,215	899	936
	Estimated cost of protection.....						49

† Estimate based upon ratio of selected to total dutiable exports to U.S.

of Canadian exports to the United States of \$936 million. The detailed increases, which are shown in columns 5 to 7 of Table 2, consist chiefly of crude and processed agricultural products and foodstuffs, primary metal products, and transportation equipment. Again, using the previously described method for calculating the cost of protection, the estimated welfare gain from the increase in imports was \$49 million, or .01 percent of United States national income in 1961.

Judged in relation to 1961 United States imports from Canada, the estimated increase was 77 percent of dutiable imports and 30 percent of total imports. In relation to total Canadian exports, the estimated increase was 16 percent. While the estimated increases do not appear unusually large, it is nevertheless of interest to consider whether Canadian manufacturing industry, in particular, could in fact expand its exports to the United States by the amounts indicated. The literature on Canadian-American economic relations [1] [4] suggests certainly that many branches of Canadian industry would be able to do so, especially because of the belief that the American tariff has fostered Canada's relative cost disadvantages by restricting its secondary industries to markets of inadequate size.<sup>17</sup>

Indeed, a comparison of the relative per man-hour wages and of the value added per man-hour in the major manufacturing industries in the two countries for 1960 does reveal pronounced relative cost disadvantages for Canada in a number of industries.<sup>18</sup> However, the extent to which this may be due to the American tariff is unclear, since much of Canadian industry is itself protected by tariffs. A unilateral reduction of American tariffs might not call forth substantial increases in imports from Canada, therefore, unless Canadian producers were actually able to reduce costs significantly with the expansion of the market for their goods.

What are the implications of a joint removal of tariffs by the two countries? In such an event, the facts suggest that a substantial proportion of Canadian secondary industry would face difficult adjustment problems in the short run. The nature of these problems and the effects which free trade between the two countries might have upon the structure of industry and the location of production could be determined of course only by detailed industry studies.

Some insights into the effects of free trade might possibly be sought in the experience of the Canadian agricultural implements industry, which was able to expand significantly its production and exports to the

<sup>17</sup> According to [1, p. 181]: "In the broadest terms, Canada's inability to develop more effectively along these lines can be attributed in substantial measure to the restraints imposed by United States commercial policy. There can be few other general propositions about the Canadian economy which have stronger roots in economic analysis and logic."

<sup>18</sup> This was the case also for 1953, according to [4, pp. 249-54 and 263-65].

United States after exposure to American competition with the lifting of the Canadian tariff in 1944. However, the unusual circumstances surrounding the industry perhaps limit the generalizations that can be drawn.<sup>19</sup> Nevertheless, if the assertions are correct concerning the inhibiting effect of the American tariff, a number of other Canadian industries might well be able to expand and prosper under conditions of free Canadian-American trade in industrial products.

## REFERENCES

1. Irving Brecher and S. S. Riesman, *Canada-United States Economic Relations* (Royal Commission on Canada's Economic Prospects, July, 1957).
2. W. M. Corden, "The Calculation of the Cost of Protection," *Econ. Record*, May, 1957.
3. Edward F. Denison, *The Sources of Economic Growth in the United States and the Alternatives Before Us*, Supplementary Paper No. 13 (C.E.D., 1962).
4. D. H. Fullerton and H. A. Hampson, *Canadian Secondary Manufacturing Industry* (Royal Commission on Canada's Economic Prospects, May, 1957).
5. Harry G. Johnson, "The Cost of Protection and the Scientific Tariff," *J.P.E.*, Aug., 1960.
6. Guy H. Orcutt, "Measurement of Price Elasticities in International Trade," *Rev. of Econ. and Statis.*, May, 1950.
7. Howard S. Piquet, *Aid, Trade, and the Tariff* (Thomas Y. Crowell, 1953).
8. ———, *The U. S. Trade Expansion Act of 1962: How Will It Affect Canadian-American Trade?* (Canadian-American Committee, 1963).
9. Walter S. Salant and Beatrice N. Vaccara, *Import Liberalization and Employment* (Brookings Institution, 1961).
10. Frances K. Topping, *Comparative Tariffs and Trade: The United States and the European Common Market*, Supplementary Paper No. 14 (C.E.D., 1963).
11. U.S. Bureau of the Census, *U. S. Exports and Imports as Related to Output, 1961 and 1960*, Series ES 2 No. 4 (Bureau of the Census, 1963).
12. U.S. Bureau of the Census, *United States Imports of Merchandise for Consumption*, Report No. FT 110 (Bureau of the Census).
13. U.S. House of Representatives, Subcommittee on Foreign Trade Policy of the Committee on Ways and Means, *Foreign Trade Policy: Compendium of Papers on United States Foreign Trade Policy* (U.S. G.P.O., 1958).
14. Beatrice N. Vaccara, *Employment and Output in Protected Manufacturing Industries* (Brookings Institution, 1960).

<sup>19</sup> Most notably, there was free entry into the American market after 1913, but a protected market for Canadian producers until 1944. The very favorable market conditions immediately following World War II were also conducive to expansion of the production and trade of the industry.

## DISCUSSION

LAWRENCE B. KRAUSE: For people delivering papers at these meetings, optimum strategy demands that they present just enough substance in their remarks to be interesting, but not enough to dissuade the profession from reading their future reports of research. This strategy also has many advantages from the point of view of discussants since they can point to errors of omission—always more cheerfully received than errors of commission—while attributing the author's shortcomings to space limitations. Sessions can then be ended with exchanges of mutual congratulations. Such a ritual should be preserved and Professors Eastman and Stern have done their part most admirably.

We have been provided with some delicious morsels, but after having digested them, one's appetite is not completely satisfied. Professor Eastman approached his task quite differently from Professor Stern. Eastman has concentrated his attention on the effects of protection upon the organization and behavior of Canadian industries, while Professor Stern has dealt exclusively with the effects of U.S. protection on international specialization. If our participants had to choose one or the other approach, one might have guessed that the decisions would have been reversed. Since Canada has a relatively small economy, removal of protection might bring complete specialization in certain lines of activity with substantial relative shifts between expanding and contracting industries. The United States, on the other hand, is not only large, it has substantial natural protection from transportation costs, etc.; and one would expect that the removal of artificial barriers to trade might result in very little relative shifting of production abroad. However, there is a great deal of interest as to the effects of tariff reduction on the conditions of competition within U.S. industry. While the papers by Eastman and Stern are certainly not uninteresting, I do think that a reversal of the approaches would be equally interesting.

Let me now consider Professor Eastman's paper in detail. Most of his analysis was addressed to a chain of reasoning which might briefly be summarized as follows: The existence of tariffs in Canada permit high domestic prices, which in turn allows inefficiency to be maintained; this inefficiency results primarily from the smallness of Canadian markets relative to optimum minimum plant size, and furthermore, a lowering of domestic prices would reduce inefficiency and, therefore, a lowering of tariffs would be desirable since it would tend to reduce domestic prices. I wish to examine a number of links in this chain:

One must begin with a short comment on the unit of measurement of inefficiency used by Eastman. He has estimated the proportion of individual pieces of equipment or of plants that were of minimum optimum size and over. As Eastman has recognized, this measure does not distinguish between the degrees of inefficiency of that capacity of less than minimum size. What is required is a continuous variable (in money terms) and the lack of such a

measure is more significant than Eastman suggests, as I hope to show subsequently.

In an attempt to determine the cause of inefficiency, Eastman has utilized a sample of thirteen industries for which six characteristics were examined as causation variables. Because of the paucity of observations, each characteristic was examined separately and the results were assumed to be additive. But the problem really requires a multivariate analysis. By not doing this, it is possible to overlook the interactions between the independent variables themselves. In particular, transportation costs, regionality of markets and the market size-minimum optimum plant ratios are all obviously important, but since they appear to be correlated amongst themselves, some of the explanatory power attributed to the market size-minimum optimum plant ratio may in fact be due to the other two factors.

The weakest link in the line of reasoning followed by Eastman was the attributing of high prices in Canada (measured in relation to a competitive efficiency level) to the degree of inefficiency present in the industry. As Eastman points out, high prices are found in Canada in both efficient and inefficient industries, the only difference being that in the latter either excess profits or high selling costs are also present. Clearly, price setting in Canada requires more explanation and probably would turn on the question of the number of competitors and the quality of competition within industries.

Finally, turning to the issue of tariffs, tariffs are recognized as a permissive condition for inefficiency, not a cause per se. In other words, tariffs are not a sufficient condition for inefficiency and I would like to ask whether tariffs are even a necessary condition. In order to answer this question, one must examine what might happen if tariffs are removed. I am not so certain of this result. My investigation of the United States steel industry and imports did not find this expected result.<sup>1</sup> In an oligopolistic industry, imports can be considered as the equivalent to the entry of a new firm into the industry. The effect of the new firm on industry prices will depend on (1) the efficiency of the new firm (prices of the foreign, i.e., U.S. competition) and (2) the capacity of the new firm (the excess capacity of U.S. firms and the transportation costs from U.S. production points to Canadian consumption points). If the capacity of the new firm is small, the profit maximizing behavior of the industry will dictate that they maintain their price level and share the market and thus become more inefficient. If the capacity of the new firm is great, then the degree of inefficiency of Canadian industry will determine whether the industry can adjust to a lower price level or whether it will disappear completely. In essence, one cannot merely assume that a competitive response will be forthcoming to import competition when the industry under examination is known to lack competition. In summary, I have a great deal of sympathy with the conclusions reached by Eastman, but I believe that more work is required to prove the case.

Now turning to the paper by Professor Stern, the research problem given him is not an easy one to deal with. Every economic variable that is rele-

<sup>1</sup> Lawrence B. Krause, "Import Discipline: The Case of the United States Steel Industry," *J. of Ind. Econ.*, Vol. XI, No. I.

vant for questions concerning domestic production and consumption, plus all of the special characteristics of international trade, are of some importance to the problem at hand. Furthermore, the minute detail classification of the U.S. tariff schedule containing close to 6,000 separate items complicates the problem all the more. In order to make any headway in analysis, some simplification of the problem is required. Stern has chosen to drastically reduce the number of items selected for investigation by concentrating on the products previously investigated by Piquet, he has arbitrarily assigned domestic demand and domestic supply elasticities to these products, and he has assumed perfectly elastic foreign supply curves for each product. As Professor Stern has recognized, these assumptions can lead to substantial errors in the results.

By using these assumptions, Professor Stern was able to estimate that if all U.S. tariffs were eliminated, our imports would increase by \$2.6 billion based on a 1951 level of trade and using Piquet's estimate of the effects of quota suspension, Stern suggests that imports would rise by \$3.4 billion with full liberalization. Relating these magnitudes to GNP and evaluating their welfare implications, Stern concludes that protection is maintained at relatively small cost to the U.S. economy and adjustment to import liberalization could be easily accomplished. If one takes Professor Stern's estimate of trade effects as being approximately correct and if liberalization had no further consequences for the economy, one might well conclude on the contrary that all of the efforts required to bring about trade liberalization are misdirected given their very small return.

However, I would deny that the cost of protection is so small for the United States and that adjustment to free trade would be painless. In the first place, the estimated import increases by product class may be misleading because the interrelatedness of the tariff changes have not been explicitly recognized. A tariff reduction on woolen textiles, for instance, may appear very damaging to the American textile industry when viewed in isolation, but may appear quite differently if simultaneously the substantial tariff on raw wool, coal tar dyes, and other inputs into the textile industry are also reduced. The aggregate increase in imports may not be as badly estimated as the distribution of that increase between products with very important implications for the U.S. economy. Furthermore, the impact on the domestic industry may not be adequately reflected by the change in the value of imports. The recent experience of the steel industry, the automobile industry, and the heavy electric generator industry is illustrative of this point. Also the longer-run implications of a tariff change may be much greater than the impact estimated by Stern. The assumed supply elasticity of zero for crude materials, for instance, could only be approximately correct for the very short run. Finally, the products excluded from the study may be of great significance if the current level of tariffs provides absolute protection for these goods.

As for the estimate of import increases that would result from a suspension of all quotas, I would venture to say that the estimate made by Piquet in 1953 is no longer usable even as a guide to relative magnitude. Piquet's



estimate did not adequately reflect the implications of the undermining of all domestic agricultural price support programs that would result from quota elimination, nor does the adjustment made to have it reflect the many increases in quotas we have enacted since the early 1950's appear sufficient. The quotas on imports of lead and zinc, on imports of petroleum, and the so-called "voluntary restrictions" of Japanese exports of many products to the United States were all developed after Piquet made his estimate and are very important. The implication of their removal would be very great indeed and would have to be investigated in great detail before a meaningful estimate could be made.

Turning to the question of the difficulty of adjustment to trade liberalization for the United States, the discussion by Professor Stern seems quite unrealistic since it does not explicitly recognize the state of aggregate demand and balance of payments as relevant variables. Even a modest increase in competitive imports during periods of substantial unemployment may be difficult to adjust to and its welfare implication may be negative rather than positive. Also measuring the adjustment in relation to GNP as Stern has done suggests that trade dollars are identical to GNP dollars. But this requires a flexible exchange rate which we do not have. Given the lack of a quick acting international adjustment mechanism and an inadequate international liquidity mechanism, the divergence between the two dollars could be very great. An increase of \$5 billion of imports with no feedbacks on inpayments, for instance, with our current level of unemployment and our present balance-of-payments problem would be nothing short of a catastrophe. Our policy-makers might well trade \$30 billion of GNP to avoid such a situation and think that they made a good deal. To be realistic, it must be recognized that the United States simply cannot afford a unilateral movement toward free trade and that other countries' tariff reductions put a brake on how far the U.S. can go.

As to the trading relations between Canada and the United States, very little can be said without detailed study. I understand that some efforts are now being directed toward making such an investigation. The economies of our two countries are, of course, already intimately related one to the other and, therefore, a complete freeing of trade should not be unmanageable. However, my intuition gives me little guidance on this matter. Having already overworked my crystal ball this year, I leave the speculation to Professor Stern.

C. P. KINDLEBERGER: It is important to address efficiency in discussing Canadian-American economic relations, but I confess that I have difficulty in equating efficiency and effectiveness in exporting, as implied by Professor Safarian's paper. Stern shows that U.S. tariffs limit exports by firms in Canada and that their removal would increase efficiency. Eastman provides an interesting measure of efficiency in the percentage of plants in an industry of minimum optimum size. It would have been appropriate to compare the size of plants in Canada owned by Canadian and American firms with each other, as well as with American best and average practice. Or profits in

American and Canadian firms in Canada. Or even the export records of Canadian and American firms in a given industry, as Professor Safarian seeks to do in one part of his paper. But why study by itself the extent to which American-owned firms in Canada export?

Foreign firms undertake operations in a given market because of some advantage over domestic firms, as well as because of profits higher than those available to them at home. They need more than the same rate of return as the local firms, even though this is higher than that at home, for otherwise it would pay them to invest through organized capital markets rather than undertake costly foreign operations directly. Foreign firms typically have an advantage in access to more profitable technology, to larger amounts of capital, or to markets. Where the advantage of the American subsidiary in Canada lies in its knowledge of or access to the American market, it will tend to export more than the comparable Canadian firm. This helps account for the substantial percentage of exporting firms (in Table 2 of Safarian's paper) who sell to their parent company in the United States, or to the United States market more generally. An American firm with an advantage in access to capital or in technology may export to the Commonwealth from a Canadian base—accounting for the 15 percent in Table 2 which exports exclusively to the United Kingdom and the rest of the sterling area. But if the American firm has an advantage over Canadian firms in the Canadian market, presumably in capital or technology, there is no reason why it should be expected to export, and the fact that it does not export is no reflection on its efficiency. Import-competing and domestic sector firms, that is, do not export. To study the export record of all American subsidiaries in Canada is certain to produce blanks in these two classes.

Or the matter can be put in terms of location theory. In market-oriented industries, American firms go to Canada in attraction to the Canadian market; in supply-oriented industries, in attraction to sources of supply. The latter will generally export; the former presumably will not, except possibly for Commonwealth preference. To study the export record of market-oriented American firms in Canada in search of measures of efficiency is futile.

It is entirely reasonable that the smaller U.S. subsidiaries in Canada should export less than the larger. The 133 firms out of 227 in the sample which both have less than \$5 million in assets and more than 95 percent of their sales in Canada are doubtless engaged in manufacturing for the Canadian market, based on an advantage in technology, and protected by the Canadian tariff. Their efficiency as compared with the best practice in the United States is a function less of their nationality than it is of protection, with which it is probably inversely correlated. The point is made by Table 3, where more of the smaller firms—market-oriented, tariff-protected inefficient—report higher costs than their parents in the United States, as compared with the larger, supply-oriented firms, which, in greater proportion, enjoy lower costs than their parents and export. The most recent and dramatic example is furnished by the Studebaker Company. The fact that it can survive in Canada when it cannot in the United States suggests

inefficiency buttressed by protection, an impression which could not be refuted by exports to other protected markets.

May I comment on the point whether one should expect a difference in export behavior among firms based on nationality? Safarian refers to the well-known theorem of Brecher and Riesman that two firms in the same circumstances, both rational, will behave in the same way regardless of nationality. On this showing, if American firms behave differently from Canadian firms, it must be because of noneconomic factors such as administrative convenience, ignorance, or response to public or private pressure. (In other national comparisons, of course, differences in tax laws should be taken into account, but these are not serious in the ordinary case between Canada and the United States.) I am not disposed to dismiss administrative convenience or ignorance as noneconomic. There are costs to decision making which favor rules of thumb over calculation and costs to overcoming ignorance. If we leave these aside, what about Safarian's view that there is no logical reason why a Canadian firm should be immune to these "noneconomic factors," which implies that it would be affected, and hence respond, in the same ways as an American firm.

But Canadian- and American-owned firms in Canada can always be expected to behave differently. This is not only because of the advantage of the American-owned firm which prompts it in the first place to overcome the costs of foreign operations. It is also a result of the fact that firms of different nationality almost certainly maximize within different horizons. What is noneconomic in one perspective is profit maximizing within a horizon which is different in space and possibly also in time profile. The American subsidiary will scan a different range of alternatives than its Canadian neighbor and may additionally maximize profits in the long run rather than the short. But even if both take the same view of profit through time and are equally spread through the globe, they will be under different pressures, from different forces of stockholder, labor, government, because they have different homes.

It would be interesting to investigate further which of Canadian- or American-owned firms in the same industry in Canada exported the more. In the export sector, I would expect that some subsidiaries had an advantage in their knowledge of or access to the American market—selling to a parent or an affiliate, for example, and taking advantage of their inside knowledge of the parent's planned production. This is the basis for international vertical integration which makes it somewhat more economical for U.S. Steel to mine ore in Venezuela through a subsidiary than to buy it on the open market, despite the costs of operating at a distance in a foreign country. Opposed to this advantage is the possibility, which looms large in Canadian opinion, that American-owned firms may restrict exports from a low cost producing point in Canada to take into account either still lower cost production elsewhere in the world, to which it alone has access, or higher cost, but politically powerful influences in the United States, which it would be dangerous to ignore. This is maximizing in the long run rather than the short. On *a priori* grounds and casual empiricism, I should have

thought that the first of these influences was more important than the second, and that American-owned firms in a given industry exported more than Canadian. Professor Safarian's limited excursion into the Canadian-American comparison suggests that the differences I would have expected are either not large or are offsetting. But here is the question that should be pushed, not the study of American-owned subsidiaries by themselves, and without division into exporting, domestic, and import-competing.

JOHN H. YOUNG: I shall offer some brief comments on the papers of Professor Eastman and Professor Stern. Professor Eastman's paper is a report on some research in a relatively neglected area of study among those concerned with the Canadian tariff. Others who have studied protected secondary manufacturing industries in Canada have observed, as Professor Eastman has, the limited scale of production of many individual products. They have gone on to argue that the primary explanation of shorter runs and larger numbers of lines produced in a single plant is the existence of the American and Canadian tariffs. There has been a tendency for some of us to think in terms of two alternatives: the situation in a particular industry as it now exists and the situation which would exist if the products of that industry moved under a free-both-ways arrangement with the United States, or, more generally, moved under a free-all-ways arrangement with the world. Less attention has been given to the changes in industrial structure which might be brought about by partial or unilateral reductions in the Canadian tariff.

This is in effect the area in which Professor Eastman has been working, testing hypotheses developed in an earlier Eastman-Stykholt article. Why is there so much small-scale, high-cost production in Canadian secondary manufacturing in spite of the fact that the size of the Canadian market is large enough to permit large-scale, low-cost production in many industries? This is a difficult question, and Eastman does not claim to have answered it. He does, however, put forward some results from a study of a sample of thirteen industries and describes these results as suggestive of the effect of the tariff on industrial structure. The sample of industries chosen contains some that do not appear to be well suited for this type of analysis. Newsprint is a leading export industry living in a free-both-ways world, cement is primarily protected by transportation costs, and one does not think of fruit and vegetable canning or beef and pork packing as typical protected manufacturing industries. For these industries, however, he finds a positive relationship between efficiency, defined in terms of the share of equipment and plant of optimum size, and the ratio of market size to minimum optimum plant. The further evidence that capital requirements act as a bar to entry is closely related to the evidence used in testing the main hypothesis, although, as Eastman points out, there may be some independent effect.

I shall not attempt to comment on the analysis of costs, profits, and prices which comes at the end of the paper. These conclusions cannot be evaluated without access to a more detailed discussion of the methods and data used in reaching them. On the whole, I thought that this paper was

well designed for a meeting of this kind. It provides some advance notice of forthcoming work and should encourage a number of us to give more thought and study to an interesting and important area of inquiry.

Professor Stern's paper can perhaps best be regarded as another in the series of papers published in recent years in which the authors argue that fundamental changes in the commercial policies of the United States, the United Kingdom, or the countries of continental Western Europe will have negligible effects on their national incomes. I shall raise one or two questions arising directly out of Professor Stern's paper and then offer a few general comments on this whole body of literature.

The first question concerns the time period to which the assumed elasticities in this paper refer. It is pointed out that the Piquet estimates might be expected to be on the low side because they are based on a temporary suspension of tariffs and quotas. It is not clear to me whether the suspension of tariffs to which Stern's assumed elasticities refer is temporary or permanent. If temporary, then presumably Stern would argue that the elasticities would be higher for a permanent change. If the assumed suspension is permanent, then leaving aside the quality of the numbers, some of the combinations and comparisons he makes are of doubtful validity.

A second query concerns the conclusion of Part I of the paper that "import liberalization could be achieved without serious adjustments in the economy as a whole." He has pointed out in footnote 11 that the removal of quotas implies discontinuance of the associated domestic programs. What is in effect being argued, therefore, is that most, if not all, of the agricultural support program, and most of the special arrangements for oil, some of the metals and other commodities could be scrapped without "serious adjustments." It is worth emphasizing that the free movement of a relatively small amount of imports may have a considerable effect on the domestic policies of government and this may also extend to the domestic policies of business and labor. Changes in these policies may in some cases have considerable short-run effects on income distribution as prices and other conditions of sale are altered to prevent a trickle of imports from becoming a flood.

In making these observations, it may be felt that I have gone over to the side of the enemy. The principal purpose of exercises of this kind is to provide assurances to the American public at large that even as radical a change as the adoption of free trade would not lead to major adjustments within the United States. It is highly unlikely that such assurances carry any weight with those in protected industries, but at least some of the statements emanating from these sources can be shown to be false. But as Professor Stern recognizes, results of this kind cut two ways. If the adjustments are small, the gains are small, and the cause of freer trade is not served by estimates which may err on the low side because of the difficulty of quantifying some of the effects, particularly the effects arising from consequential changes in the commercial policies of other countries.

This is not to say that quantitative estimates of the effects of policy changes are of limited value. Far from it. I have always been surprised at

the limited interest of the classical and neoclassical economists in the quantitative significance of the things they were talking about. Presumably our interest in these matters has been encouraged by the availability of improved statistical information and this is a very healthy development. We are continually being subjected to predictions of disaster expected to result from automation or peace or commercial policy developments in other parts of the world, and alternatively are promised economic bliss if we will only adopt this or that line of policy. It is part of the function of the professional economist to subject these propositions to historical and quantitative analysis and reveal the lack of perspective. At the same time, it behooves economists to be properly modest about the contribution our profession can make to economic well-being. On the face of it, it is curious that at a time when great efforts are being made to apply economizing principles to the allocation of defense expenditure or to the use of resources within firms, economists should be denigrating the importance of these principles in the operation of entire economies. The explanation in part is probably that when working in some of these narrower areas, economists need only show that they are more than earning their keep. We perhaps need something of the same attitude in thinking about national economic policy. Economists alone are not responsible for keeping gross national product from falling to zero, nor are they capable of providing ideas which will double gross national product in the near future. Indeed, there is probably a fairly narrow range of national income over which academic economists and their brethren in government and elsewhere can exercise much influence. In the short run, this is much wider in the area of monetary, debt, fiscal and exchange rate policies than in policies primarily affecting the allocation of resources; but in the long run it may be fairly narrow for both. If it were possible to estimate this range, it might be more sensible to relate the economic gains of a proposed policy change to the total economic gains likely to be achieved by the general adoption of judicious economic policies. Using this method of calculation it might turn out that the adoption of free trade was by no means unimportant even for the United States.

# EFFICIENCY OF THE SOVIET ECONOMY

## THE STATIC EFFICIENCY OF THE SOVIET ECONOMY

By JOSEPH S. BERLINER  
*Brandeis University*

In the discussion of economic efficiency, all roads lead back to Pareto. The optimality criterion directs us to seek for evidence that permits us to judge whether the bundle of goods produced is such that no other attainable bundle could contain more of some goods and no less of any other good. Two types of evidence can be marshaled. The first is theoretical in character and is based on the conditions that must be satisfied if an optimal bundle is to be obtained. With no empirical data at all, for example, if we knew that the methods of price formation were such that prices could not be expected to correspond to marginal rates of transformation, we could conclude that insofar as prices governed resource allocation, that allocation could not be optimal. The second method is based on empirical data. If we found evidence that the system tended to leave productive resources unused, or used in less productive employments, or that output was produced which could not be sold, we could again conclude that the allocation was less than optimal.

A great deal of evidence can now be drawn upon for analyzing sources of inefficiency in the Soviet economy from both the theoretical and empirical points of view. Studies of the methods of price formation have directed our attention to the consequences of the use of average instead of marginal costs, of failure to account for interest and obsolescence (until very recently), of the complex system of multiple pricing, and so forth. We have also a considerable number of empirical studies of price and cost relations, of the consequences of physical planning with material balances, and of the results of the system of success indicators, and so on. It would be a feasible approach to my topic to marshal all the evidence of inefficiency and present it as a sort of tally sheet. The trouble, however, is that there is no evident way of actually tallying it up. And short of a technique for tallying, it would remain nothing more than an anecdotal account, even though quantified in parts. However suitable such a technique might be for presenting a legal brief or a political tract, the amassing of pieces of anecdotal evidence is no substitute for a summary calculation. One wild price would serve as well as a thousand to establish that the system was less than

perfectly efficient, but neither one nor a thousand would answer the question, "How inefficient?"

For I take it as evident that no real economic system is perfectly efficient. It is indeed of interest to analyze the peculiar sources of inefficiency in this or that economy, but such analysis can provide no summary evaluation of the performance of a system of economic organization. Rather than deal with the sources of inefficiency of the Soviet economy, I regard my task to be an evaluation of the overall efficiency of the system.

It would be possible to confine this task to the Soviet economy alone, but the results would be difficult to interpret. Suppose, for instance, we knew of a certain mechanical system that its efficiency was 20 percent. Does it represent satisfactory performance or unsatisfactory? The answer will depend on the kind of machine and the efficiency of other machines designed for the same purpose. If it were a diesel engine, we would say it was a rather poor machine, for diesel engines have been developed with an efficiency of 40 percent. If it were a thermoelectric generator, we would say that it was remarkably good, for existing generators operate in the range of 1 to 10 percent. In like fashion, in order to interpret the significance of any indicator of the efficiency of the Soviet economy, we should have to have some notion of what is par for the economic course in general. For this purpose, I have selected the United States to serve as a standard of comparison.

One might be able to adapt a Paretian approach to a comparative study of efficiency, but there are certain difficulties. Pareto was, after all, an engineer, and it is not difficult to recognize the parentage of the production-possibilities curve in the principle of the conservation of energy. Indeed, all engineering efficiency formulae are based on the notion that there is some absolute maximum quantum of output that can ideally be got out of a system. But the maximum output that can be got out of an economic system can only be established with respect to the assumptions one might wish to make about the ideal economic system, or indeed the ideal social system. For example, the production-possibilities curve would have one position if we assumed a highly centralized planning system; another if we assumed a decentralized one; it would have one position if we took the present managerial incentive system as given; another if we assumed a different one. It would vary according to whether we assumed a better pricing system or the present one, greater freedom for private arbitrage activity or less, enlargement of the private household sector of agriculture or curtailment, and so forth. Depending on the assumptions one is prepared to make about the ideal system, the given system would turn out to have a number of different indexes of efficiency. It would, moreover, be im-



possible to compare the efficiency of different types of economic systems, for the efficiency of each would be measured relative to its own ideal, and the ideals are not likely to be commensurable.

There is one final problem with an ideal-output approach. One might find two economies perched precisely on their production possibilities curves, indicating that both were perfectly efficient in terms of their respective possibilities; and yet Economy I may produce a greater volume of output than Economy II for equivalent inputs. One might wish to interpret this as indicating that Economy II is getting as much out of its inputs as it can be expected to get under its present form of organization, but that if it were to change its organization to that of Economy I, it might get a greater output for the same inputs. The analysis of this case would require us to go well beyond the Paretian concept of efficiency.

Since the production-possibilities approach is so protean to the grasp, I have decided to reject it in favor of a different attack on productive efficiency. The situation is this: An economy at (time or place) *I* produces a certain output  $P_I$  with certain inputs. At (time or place) *II* it produces a different output  $P_{II}$  with different inputs. The question to be asked is, what would the output at *II* have been ( $\bar{P}_{II}$ ) had it been produced with the same efficiency as at *I* but with the inputs of *II*? The ratio of  $\bar{P}_{II}$  to  $P_{II}$  will be our index of the relative efficiency of *II* compared to *I*. It is only productive efficiency that is to be measured here; the problem of allocative efficiency will be conspicuously ignored.

Before proceeding with the analysis, it would be well to turn to the data in Table 1. Value data are at 1955 prices, converted into rubles or dollars by 1955 conversion ratios. The capital stock figures are gross; that is, they constitute the 1955 cost of replacing the 1960 stock of physical assets by new but otherwise identical units. The U.S. labor force data include employed and unemployed, while the Soviet figures are average annual employment. Involuntary Soviet unemployment is probably not large enough to make much difference (although involuntary Soviet employment may make a difference from a welfare point of view). The figures in parentheses refer to the total economy; the others to the nonfarm economy. Sources and details are discussed in the appendix.

Section B of the table presents the percentage relations between the U.S.S.R. and the U.S. for corresponding items. Because of the index number effect, the Soviet value magnitudes are larger relative to the U.S. when both are valued at dollar prices than when ruble prices are used. For the nonfarm GNP, the Soviet percentage is less than for total GNP, reflecting the relatively large proportion of Soviet output originating in agriculture.

TABLE 1  
OUTPUTS AND LABOR AND CAPITAL INPUTS, 1960

A. BASIC DATA				
	Original Prices		Converted Prices	
	U.S.S.R.	U.S.	U.S.S.R.	U.S.
1. GNP (billion rubles or dollars).....	1,300 R (1,730 R)	\$423 (\$441)	\$226 (\$300)	4,870 R (5,080 R)
2. Capital stock (billion rubles or dollars).....	2,546 R (2,965 R)	\$1,400 (\$1,513)	\$509 (\$593)	9,800 R (10,590 R)
3. Labor (million man-years).....	57.9 (95.7)	64.8 (70.7)	57.9 (95.7)	64.8 (70.7)
B. U.S.S.R. AS % OF U.S.				
	At Dollar Values		At Ruble Values	
4. GNP.....	53.5 (68.)		26.6 (34.1)	
5. Capital.....	36.4 (39.1)		26.0 (29.7)	
6. Labor.....	89.3 (134)		89.3 (134)	
7. GNP per unit of labor.....	59.9 (50.7)		29.8 (25.2)	
8. GNP per unit of capital.....	147.0 (174.0)		102.2 (115.0)	

SOURCES AND EXPLANATIONS: See appendix.

NOTES: Figures in parentheses refer to the total economy. The others refer to the nonfarm economy.

Soviet GNP at original prices is gross of turnover taxes and subsidies. An attempt was made to eliminate the effect of turnover taxes and subsidies in the converted values. See appendix.

The difference between the nonfarm and the total capital stocks is not as great as the difference between the nonfarm and total GNP. This reflects the relatively less capital intensive organization of Soviet agriculture. The total Soviet labor force is 34 percent greater than that of the U.S., but the nonfarm labor force is only 89.3 percent of that of the U.S., again reflecting the relative labor intensity of Soviet agriculture.

Rows 7 and 8 of Table 1 present the single-factor productivity ratios implied in the basic data. Soviet labor productivity (row 7) is in all cases less than that of the U.S., although in the nonfarm sector it is relatively higher as one would expect. The capital productivity figures reverse the picture (row 8). Since the Soviet capital stock falls short of the U.S. by more than Soviet output falls short of U.S. output, Soviet capital productivity is higher than that of the U.S. The Soviets get 47 percent (74 percent) greater output per unit of capital than the U.S. in dollar terms and 2 percent (15 percent) more in ruble terms.

The nature of the task of computing relative efficiency is now clear. The U.S.S.R. produces a smaller output than the U.S., but it does so by using considerably less capital and somewhat less labor (in the nonfarm economy; in the total economy it uses more labor). The relative

efficiency of the U.S.S.R. therefore depends on the weights assigned to labor and capital. The choice of weights, which is the crucial decision, reflects the assumption one makes about the shape of the production functions.

Using John Kendrick's [5] method for our first set of computations, the weights are the relative factor shares of labor and capital. By assuming that the factors are paid the value of their marginal product and that these factor payments exhaust the product, Kendrick requires his production function to be homogeneous of degree one. By using constant product and factor prices, the production isoquants and the product transformation functions are treated as if they were linear. Hence the production function is of the form  $P = aL + bK$ . The relative efficiency of Economy II to Economy I is then given by the formula:

$$\frac{P_{II}/P_I}{(L_{II}/L_I) \cdot \alpha_I + (K_{II}/K_I)(1 - \alpha_I)}$$

where  $\alpha_I$  and  $1 - \alpha_I$  are the factor shares of labor and capital in Economy I.

In adapting the formula to our purposes, one must heed Moorsteen's [6] caution that "the input index employed clearly must refer to inputs that are both sufficient and necessary to produce the outputs measured by the output index" (p. 461). This condition is satisfied if the output index uses the prices of one country and the input index uses the prices of the other. (See Moorsteen for the proof.) Applying this condition to Kendrick's formula yields two indexes of relative Soviet efficiency:<sup>1</sup>  $E_1 = 33.8$  percent (29.8 percent); and  $E_2 = 64.1$  percent (60.2 percent) with Soviet returns to capital at 8 percent, and 90.0 percent (73.7 percent) with Soviet returns to capital at 20 percent (Table 2). Index  $E_1$  has the following meaning: Assume that both countries have linear isoquants with slopes proportional to U.S. factor prices, and linear transformation functions with slopes proportional to Soviet output prices; then if either country had used the inputs of the other, Soviet output would have amounted to 33.8 percent of U.S. output.  $E_2$  has a

$$^1 E_1 = \frac{\frac{\sum p_{us} q_{su}}{\sum p_{us} q_{us}}}{\left(\frac{L_{su}}{L_{us}}\right) \alpha_{us} + \left(\frac{\sum k_{su} r_{us}}{\sum k_{us} r_{us}}\right) (1 - \alpha_{us})}; \quad E_2 = \frac{\frac{\sum p_{us} q_{su}}{\sum p_{us} q_{us}}}{\left(\frac{L_{su}}{L_{us}}\right) \alpha_{su} + \left(\frac{\sum k_{su} r_{su}}{\sum k_{us} r_{su}}\right) (1 - \alpha_{su})}$$

where  $k$  refers to physical capital assets,  $r$  is the asset price,  $\alpha_{us} = .8$ ,  $\alpha_{su} = .8$  when return to Soviet capital is assumed to be 8 percent, and  $\alpha_{us} = .6$  when return to Soviet capital is 20 percent. See appendix for sources.

TABLE 2  
SOVIET PRODUCTIVE EFFICIENCY AS % OF U.S.

	Ruble-valued Outputs, Dollar-valued Inputs	Dollar-valued Outputs, Ruble-valued Inputs
$P = aL + bK$		
U.S. weights ( $\alpha = .8$ ) . . . . .	$E_1 = 33.8(29.8)$	
Soviet weights:		
$i = 8\%$ ( $\alpha = .75$ nonfarm, .8 total) . . . . .		$E_2 = 64.1(60.2)$
$i = 20\%$ ( $\alpha = .53$ nonfarm, .60 total) . . . . .		$E_2 = 90.0(73.7)$
$P = bL^\alpha K^{(1-\alpha)}$		
$\alpha = .8$ . . . . .	$E_3 = 35.6(32.3)$	$E_4 = 76.6(69.0)$
$\alpha = .7$ . . . . .	$E_3 = 39.0(36.5)$	$E_4 = 86.6(80.8)$
$\alpha = .6$ . . . . .	$E_3 = 42.6(41.4)$	$E_4 = 98.0(94.5)$
$\alpha = .5$ . . . . .	$E_3 = 46.7(46.8)$	$E_4 = 118.9(111.0)$

similar meaning, except that the input and output prices are reversed. The explanation of most of the large difference between the two results is to be found in the effect of pricing output at dollars or rubles.

The chief defect of a Kendrick-type of index for our purposes is that, like all fixed-weight indexes, it treats the underlying production and transformations functions as if they were linear. In fact the linearity assumption is only an approximation of the actual but unknown shape of the functions in the neighborhood of the observed points. But one would probably be on safer ground to assume that the production function at least is characterized by a diminishing marginal rate of substitution among factors. Perhaps the least offensive way of incorporating this property is to replace Kendrick's production function by one of the Cobb-Douglas variety,  $P = bL^\alpha K^{1-\alpha}$ . The function retains Kendrick's assumption of constant returns to scale, but now the observed inputs can slide along nice convex isoquants instead of being forced to move along straight lines. It would be pleasant if some concavity could also be impressed on the product-transformation functions, but linearity will have to continue to hold sway in that domain.

Assuming the Cobb-Douglas function, if we are given one observation on  $P$ ,  $L$  and  $K$ , then for any assumed value of  $\alpha$  the parameter  $b$  is determined and the whole production surface defined. It is then possible to predict what the output would be, for any  $\alpha$ , if a different set of inputs were used. Bearing Moorsteen's theorem in mind, we can thus construct two new indexes of efficiency, both of the form

$$\frac{P_{II}/P_I}{\left(\frac{L_{II}}{L_I}\right)^{\alpha I} \left(\frac{K_{II}}{K_I}\right)^{(1-\alpha I)}}.$$

$E_3$  uses Soviet price weights in the numerator and U.S. factor prices in the denominator, while  $E_4$  uses the prices in reverse.<sup>2</sup> The efficiency indexes have been computed, not only for weights corresponding to factor shares in the two countries, but for a series of values of  $\alpha$  from .8 to .5. The indexes are to be interpreted as indicating what percentage Soviet output would be of U.S. output if either country had the indicated value of  $\alpha$  and used the other country's quantities of inputs.

The results, in Table 2, indicate that for the same  $\alpha$  the Cobb-Douglas index yields a somewhat higher value for Soviet relative efficiency than does the Kendrick-type index. With an  $\alpha$  of .8, the Kendrick is 33.8 percent, the Cobb-Douglas 35.6 percent. With an  $\alpha$  of .6, the two indexes (for the total economy) are 73.7 percent and 94.5 percent. Note that Soviet relative efficiency increases as the value of  $\alpha$  declines (or as the capital exponent increases). The reason is that the use of U.S. inputs by the U.S.S.R. is equivalent to the substitution of capital for labor. The greater the contribution of capital to production, the more the Soviets benefit by undertaking the substitution (and the more the U.S. would lose by shifting to Soviet inputs).

Our judgment about Soviet relative efficiency depends then on what the value of  $\alpha$  actually is. Randomly collected estimates for various times and places confront one with a bewildering variety: .76 for the U.S. in 1919, .64 for Australia in 1934-36, and .43 in Canada in 1937 (Douglas); .65 for the U.S. (Solow); .40 for India (Divatia and Trivadi); .20 for the U.S. (Valavanis); and so forth. It is difficult, however, to accept values of less than .5 for the U.S.S.R. It would surely require an increase of more than 10 percent in the Soviet capital stock to compensate for a reduction of 10 percent in the labor force. If one had to stab at a more or less reasonable set of values, they might lie in the range of .6-.7 for the U.S.S.R. and .7-.8 for the U.S. At these values, the relative efficiency of the U.S.S.R. ranges from 36-39 percent of the U.S. if the U.S. used Soviet inputs and if outputs were valued at Soviet prices, to 87-98 percent of the U.S. if the U.S.S.R. used U.S. inputs and if outputs were valued at U.S. prices.

Because of the index number problem there is probably not very much one can do to narrow the range of these indexes, short of cutting Gordian knots by splitting differences. But there is a great deal that could be done to narrow the range of items that contribute to the

$$E_3 = \frac{\frac{\sum p_{su} q_{su}}{\sum p_{su} q_{us}}}{\left(\frac{L_{su}}{L_{us}}\right)^{\alpha_{us}} \left(\frac{\sum k_{su} r_{su}}{\sum k_{us} r_{us}}\right)^{(1-\alpha_{us})}}; \quad E_4 = \frac{\frac{\sum p_{us} q_{su}}{\sum p_{us} q_{us}}}{\left(\frac{L_{us}}{L_{su}}\right)^{\alpha_{su}} \left(\frac{\sum k_{su} r_{su}}{\sum k_{us} r_{us}}\right)^{(1-\alpha_{su})}}$$

where the symbols have the same meaning and values as in the preceding footnote.

measured differences in efficiency. A substantial proportion of what has been registered here as relative Soviet inefficiency would be accounted for by the fact that the factors of production have not actually been held constant in all relevant and possible ways. Land has not been included at all, and a certain portion of what has been registered as U.S. efficiency may be credited simply to a better natural endowment. Similarly, the factors are not homogeneous. The labor factor in the two countries is not homogeneous with respect to skills, hours of work per year, and other relevant respects. Capital is not homogeneous with respect to average age, wear and tear, or intensity of utilization.

Since factor differences are included in what is measured by the relative efficiency indexes computed above, they should perhaps be referred to as "gross" indexes. They may be regarded as measuring the relative efficiency of total national economies rather than the efficiency of pure systems of economic organizations. They tell us something about the gross result of the production efforts of two nations: the Soviets with all their poor agricultural resources and so forth; the U.S. with all its unemployed and so forth. This is interesting to know. But the gross index can yield no conclusion about the relative efficiency, for example, of Soviet socialism and American capitalism as pure systems of resource allocation. One could approach an answer to this question, however, by cleaning up the data to render them more homogeneous and therefore capable of yielding a "net" index of relative efficiency.

## APPENDIX

### *Gross National Product*

U.S.S.R.: Bornstein's [2, p. 385] estimate of U.S.S.R. GNP at established prices in 1955 was 1,285.8 billion rubles. Assuming a 6 percent rate of growth during 1955-60, the 1960 figure at 1955 prices would be 1,730 billion rubles. Bornstein estimated that 27.1 percent of national income originated in agriculture in 1955. I assume that the figure fell somewhat by 1960, to approximately 25 percent, yielding a nonfarm output of about 1,300 billion rubles.

Bornstein's 1955 GNP of 1,285.8 billion rubles converted to a dollar figure of \$212.4 billion, implying a ruble/dollar ratio of about 6. Applying this ratio to the 1960 estimates yields \$288 billion for total GNP and \$217 billion for nonfarm GNP.

Since established Soviet prices are gross of turnover taxes and subsidies, they understate the dollar value of Soviet output relative to the U.S. Bergson [1, p. 295, fn. 9] estimated the extent of the understatement to be about 4 percent. Adjusting the converted figures by 4 percent increases them from \$288 billion and \$217 billion to \$300 billion and \$226 billion.

U.S.: The 1960 GNP at constant 1954 prices was \$440.8 billion [7, p. 314]. The wholesale price index was virtually unchanged from 1954 to 1955. National income originating in agriculture in 1960 at 1954 prices was 4.1

percent of total national income. Applying that percentage to the GNP figure above yields a nonfarm GNP of \$422.7 billion.

Bornstein's [2] 1955 U.S. GNP figure of \$397.5 billion converts to 4,802 billion rubles, implying a ruble/dollar ratio of about 12. Applying this conversion factor to the above results yields the figures of 5,290 billion rubles and 5,072 billion rubles.

Applying Bergson's correction factor of about 4 percent to compensate for turnover taxes and subsidies in Soviet prices, the above figures are reduced to 5,080 billion rubles and 4,870 billion rubles.

### *Capital Stock*

U.S.S.R.: In 1959 the U.S.S.R. undertook a general capital inventory and revaluation. The results, as analyzed by Kaplan [4] are a Jan. 1, 1960, total of 2,964.7 billion rubles at 1955 prices. The figure is gross of both wear and tear and depreciation; i.e., it represents the 1955 cost of replacing the 1960 stock of capital in place by new assets of the original type. It includes housing and livestock.

The agricultural component is reported to be 419.2 billion rubles, leaving a nonfarm total of 2,546 billion rubles.

Bornstein's [2] "very rough" ruble/dollar ratio for investment end-items, with Soviet weights, is 5. Applying this to the ruble capital stock figure yields \$593 billion for the total and \$509 for nonfarm.

U.S.: The basic figures are from Goldsmith [3]. For a coverage equivalent to that of the Soviet figure, I added structures, producers durables and livestock, for a total of \$1,257.8 billion and a nonfarm component of \$1,162.2 billion (pp. 123, 196-200). The figures refer to Jan. 1, 1959, and are valued at the prices of 1947-49. I estimate 1959 gross investment less retirements to be \$36.5 billion total and \$34.3 billion nonfarm. Applying the wholesale price index for 1955 of 117 (1947-49 = 100) yields a final estimate of \$1,513 total and \$1,400 nonfarm, for the Jan. 1, 1960, capital stock at 1955 prices.

The figures are taken from Goldsmith's "gross" capital stock series, which were compiled by the "perpetual inventory" method; i.e., gross investment less retirements. Like the Soviet figures, they are gross of both wear and tear and depreciation. They represent the 1955 cost of reproducing the 1960 capital stock in place with new assets equivalent to the original ones.

Bornstein's ruble/dollar ratio for investment at U.S. weights is 7. Applying this to the dollar figures yields ruble figures of 10,591 billion rubles total and 9,800 billion rubles nonfarm.

### *Labor*

U.S.S.R.: The figures are those compiled by Weitzman, Feshbach and Kulchitsky [9, p. 649]. The figures correspond closely to average annual employment in man-years, and were designed for comparability with U.S. data.

U.S.: Weitzman [9] has also recomputed U.S. employment data to make them comparable to Soviet data. Weitzman's figures have been adjusted

here in order to add the unemployed to the U.S. data. The BLS unemployment data [8, p. 57] of 5.1 percent in nonagriculture and 3.0 percent in agriculture were applied to Weitzman's "constructed series" to obtain the final estimates.

### *Input Weights*

U.S.S.R.: Bergson [1, p. 422] gives the average annual wage in 1955 as 8,520 R. *Vestnik statistiki*, 1961, No. 5, reports that average money earnings of wageworkers in industry rose 17 percent from 1955 to 1960. I therefore assume the 1960 average wage to have been around 10,000 R. Applying this to the nonagricultural labor force of 57.9 million gives a total wage bill of 579 billion R.

I assume the real income in agriculture, including income in kind, to have been about 8,000 R. Applying this to the agricultural labor force of 37.8 million, gives a money income of 302 billion R.

Returns to capital are obtained by applying the arbitrary figures of 8 percent and 20 percent to the stock of capital in Table 1. The following are the resulting proportions of income going to labor and capital respectively: at 8 percent return to capital, .74 and .26 for nonfarm, .79 and .21 for total; at 20 percent, .53 and .47 for nonfarm, .60 and .40 for total.

U.S.: Kendrick [5, p. 285] gives .79 and .21 for labor and capital shares, during 1948-53.

### REFERENCES

1. A. Bergson, *The Real National Income of Soviet Russia Since 1928* (Harvard Univ. Press, 1961).
2. Morris Bornstein, "A Comparison of Soviet and United States National Product," in Joint Economic Committee, U.S. Congress, *Comparisons of the United States and Soviet Economies* (1959), Part II.
3. R. W. Goldsmith, *The National Wealth of the United States in the Postwar Period* (Princeton Univ. Press, 1962).
4. Norman M. Kaplan, "Capital Stock," in A. Bergson and S. Kuznets (eds.), *Economic Trends in the Soviet Union* (Harvard Press, 1963).
5. John W. Kendrick, *Productivity Trends in the United States* (Princeton Univ. Press, 1961).
6. Richard H. Moorsteen, "On Measuring Potential and Relative Efficiency," *Q.J.E.*, Aug., 1961.
7. Bureau of the Census, *Statistical Abstract of the United States, 1962* (GPO, Washington, D.C.).
8. U.S. Bureau of Labor Statistics, *Employment and Earnings*, Annual Supplement, June, 1962.
9. M. Weitzman, M. Feshbach, and L. Kulchyska, "Employment in the U.S.S.R.," in Joint Economic Committee, U.S. Congress, *Dimensions of Soviet Economic Power* (GPO, 1962).



# THE DYNAMIC EFFICIENCY OF THE SOVIET ECONOMY

By BELA BALASSA

*Yale University*

"A system—any system, economic or other—that at every point of time fully utilizes its possibilities to the best advantage may yet in the long run be inferior to the system that does so at *no* given point of time, because the latter's failure to do so may be a condition for the level or speed of long-run performance."—JOSEPH A. SCHUMPETER.

## I

Professor Bergson summed up the debate of the interwar period on socialism stating that the controversy boils down to the question: "Which is more efficient, socialism or capitalism?" [5, p. 447]. But what is meant by efficiency? A perusal of the contributions to the debate suggests that the discussion centered around static efficiency; i.e., efficiency in the allocation of resources among competing ends at a point of time. I have elsewhere argued that static efficiency is only one of several possible "success indicators" that can be used to evaluate the workings of economic systems. Other suggested indicators are dynamic efficiency, the growth rate of national income, consumer satisfaction, and the distribution of income. Dynamic efficiency has been defined in relative terms, with reference to a comparison of economic systems: assuming identical initial resources (including labor) and identical saving ratios, a higher rate of economic growth would correspond to a higher degree of dynamic efficiency [2, p. 10].

While static efficiency in production means that the society operates on its production-possibility frontier, dynamic efficiency can be represented by the movement of this frontier in a northeast direction. The concept was originally used by Dorfman, Samuelson, and Solow for the case when production functions are linear homogeneous and technological change does not take place. It can be shown that, under these assumptions, perpetual static (one-period) efficiency does not guarantee the attainment of dynamic (intertemporal) efficiency, while dynamic efficiency—the maximum rate of northeast movement of the production-possibility frontier compatible with technological possibilities, consumption profiles, and initial conditions—requires that the allocation of resources be efficient at every point of time. If we add the further assumption that producers know current prices and instantaneous rates of price changes, the conditions for dynamic efficiency will be fulfilled under perfect competition [9, Chap. 12].<sup>1</sup>

<sup>1</sup> The problem of the stability of competitive equilibrium is not considered here.

Different conclusions are reached if we remove the assumption of unchanged technological knowledge and constant returns to scale and also consider the effects of uncertainty on economic behavior. Now Schumpeter's dictum can be applied, according to which the fulfillment of the conditions of static efficiency is not a precondition of dynamic efficiency; inasmuch as, e.g., oligopolistic market structures may be more conducive to innovation than pure competition.

Allowing for uncertainty and the possibility of technological change, the major influences determining dynamic efficiency will include innovating activity, the spreading of new technology among the units of production, and the choice of appropriate investments. The dynamic efficiency of a national economy could then be indicated by comparing its actual growth rate to its potential growth rate, defined as the maximum rate of northeast movement of the production-possibility frontier compatible with consumption profiles and initial conditions, including the rate of saving.

We can hardly expect to be able to determine the potential rate of growth of a national economy, however, and an empirical investigation of dynamic efficiency would have to rely on intercountry comparisons instead. It can be surmised that two economies with identical initial conditions and consumption profiles would have the same growth-potentialities, and hence a comparison of their growth rates would indicate their relative dynamic efficiencies, provided that disparate changes in static efficiency do not take place.

For national economies with different initial conditions and disparate changes in static efficiency, the net effect of the influences contributing to dynamic efficiency could be ideally measured by adjusting for the contribution of identifiable inputs and changes in static efficiency to economic growth. But the comparisons have to be restricted to economies at a similar level of development, since a national economy at an early stage of industrialization would have an advantage in the possibility of applying technology developed in industrial countries. Finally, in the case of national economies that reached the same level of economic development at different points of time, a comparison of the corresponding periods will be distorted by reason of the fact that the economy industrializing in a later period could draw on a fund of accumulated knowledge that was not available to its counterpart before.

The dynamic efficiency of a national economy will thus reflect the ability of the economy to generate innovations, to disseminate technological know-how, and to appropriately allocate new investments. These activities, in turn, are affected to a considerable extent by the institutional structure of the economy. A comparison of the dynamic efficiency of national economies representative of various economic

•

systems could then give expression to the performance of these systems in regard to technological improvements and the allocation of investments.

## II

Turning to problems of measurement, our point of departure is the empirical work done in recent years on the aggregate production function. It is customary to assume that this function takes the following form:

$$(1) \quad Q = A(t)f(T_1, T_2, T_3, \dots, T_n)$$

The second part of the right-hand side of the equation is the "traditional" production function, indicating the relationship between identifiable inputs and outputs, while the term  $A(t)$  is a shift variable, expressed as a function of time. This has been called, among other things, "total factor productivity," "change in productive efficiency," "technical change," and, also, "the measure of our ignorance." The shift in the aggregate production function is usually taken to be unrelated to changes in the quantity of inputs,<sup>2</sup> and it is also assumed to raise the marginal productivities of all factors in the same proportion. Postulating the equality of factor prices and marginal products, relative shares would thus be invariant to shifts in the function.

Assumptions need further be made with regard to returns to scale and substitutability among the factors of production. Customarily, a linear homogeneous production function (constant returns to scale) and a constant elasticity of substitution among factors is assumed. Production functions with a constant elasticity of substitution can, in turn, be classified according to the assumed values of the elasticity of factor substitution—a Leontief input-output system with zero substitution elasticity, a Cobb-Douglas function with a unitary elasticity, a linear production function with infinite substitution elasticity, and, finally, a general class of CES (constant elasticity of substitution) functions.

Special interest attaches to production functions with unitary and infinite elasticity of substitution, since—explicitly or implicitly—these have been used in most empirical studies utilizing an aggregate production function. These functional forms are shown in equation (2) and (3) for the two-factor case.

$$(2) \quad Q = A(t)L^\alpha K^\beta, \text{ where } \alpha \text{ and } \beta, \text{ the relative shares of labor and capital,} \\ \text{add up to 1}$$

<sup>2</sup> By contrast, the shift in the function is assumed to be dependent on new investments in a paper by R. M. Solow [15], with gross investment providing a constraint to the introduction of technological improvements.

- (3)  $Q = A(t)(aL + bK)$ , where  $a$  and  $b$  are the marginal product of labor and capital, respectively.

Equation (2), which applies also to a more general class of production functions, has first been used by Solow [16], while Kendrick [12] employs equation (3). After appropriate transformation, the equations can also be written as

$$(4) \quad \log \frac{Q_t}{Q_0} = \log \frac{A_t}{A_0} + \alpha \log \frac{L_t}{L_0} + \beta \log \frac{K_t}{K_0}, \text{ and}$$

$$(5) \quad \frac{Q_t}{Q_0} = \frac{A_t}{A_0} \left( \alpha \frac{L_t}{L_0} + \beta \frac{K_t}{K_0} \right)$$

According to equations (4) and (5), the first method involves a geometric and the second an arithmetic combination of inputs when, in each case, relative shares are used as weights. Thus, although the economist is inclined to dismiss the linear production function with an infinite elasticity of substitution as unrealistic, for purposes of measuring total factor productivity the main difference between the two methods of estimation appears to be that the former provides a geometric and the latter an arithmetic index. Expressing the two indices in terms of rates of change, the differences between the results will be small as long as inputs are not growing at widely different rates. In fact, in present-day industrial countries, the results appear to be little influenced by the choice between the two methods.<sup>3</sup> This conclusion is not materially affected if the general form of the constant elasticity of substitution function is used in the calculations [1, p. 224].

### III

In Section I it has been suggested that a comparison of growth rates in national economies at a similar level of economic development can

<sup>3</sup> An example given by Evsey D. Domar can serve to illustrate this point. Assuming output, labor, and capital to grow at annual rates of 3.5, 1.5, and 3.0 percent and labor's and capital's share of 80 and 20 percent, respectively, the geometric index of technological change will show an annual rate of increase of 1.7 percent and the arithmetic index an average rate of growth of 1.68 percent over a period of ten years and 1.66 percent over twenty years [8, p. 601n]. The result follows from the properties of geometric and arithmetic averages. In the case of homogeneous output, the geometric index will grow at a rate higher than the arithmetic index since the geometric average of inputs is less than their arithmetic average.

The geometric index has computational advantages if a continuous series of observations is available, however. In such cases, equation (2) can be fitted to the data and the weights (relative shares) are regarded as parameters. By comparison, larger standard errors are likely to be obtained if equation (3) is used since marginal products are in general more variable than relative shares, and the alternative of utilizing equation (5) for calculating an average of year-to-year changes becomes rather cumbersome.

provide an indication of their relative dynamic efficiencies if we adjusted for disparate changes in the quantity of inputs and in static efficiency. The question arises, to what extent estimates of total factor productivity can be utilized in gauging the degree of dynamic efficiency.

For this purpose, it would be necessary to include all identifiable inputs in the aggregate production function, and to appropriately separate inputs of differing quality, such as different varieties of land. But statistical difficulties restrict the number of inputs to be considered and changes in the quality of inputs often defy measurement. Further problems arise in attempting to adjust for changes in static efficiency. In this connection, reference should be made to intersectoral differences in the marginal productivities of the factors of production.<sup>4</sup>

Measured changes in total factor productivity are affected by the reallocation of factors from low-productivity to high-productivity occupations if aggregate factor shares are used as weights. However, this movement does not entail an improvement in dynamic efficiency but it represents an adjustment from an initial disequilibrium or an increase in inputs due to, e.g., investment in human skills.

An initial disequilibrium may reflect the existence of institutional rigidities or some degree of short-run complementarity between capital and labor that restricted the movement of labor from low-productivity to high-productivity occupations and resulted in interindustry differences in the marginal productivity of labor. This phenomenon is observed in countries at an early stage of industrialization where a certain dualism between agriculture and industry exists. Interindustry shifts eliminating—or reducing—these differences are once-for-all changes that contribute to efficient resource allocation and should be regarded as improvements in static rather than dynamic efficiency.

Interindustry differences in the marginal productivity of labor may also be due to differences in skills and educational requirements. Increases in the labor force of industries requiring skilled labor will then involve investment in human capital. Adjustments in this case, as well as in the previous one, can be made by weighting employment in the various sectors by wages paid in the base year, but there is no simple way of determining to what extent intersectoral differences in marginal productivities have been due to interindustry differences in factor qualities or to inefficiencies in resource allocation.

Lack of mobility is only one of several possible causes for static inefficiencies, however. Thus, estimates of total factor productivity can give expression to improvements in static, as well as in dynamic, effi-

<sup>4</sup> The reader will note that while we deal here with the problem of intersectoral differences in marginal productivities, similar considerations apply to interregional (North-South), interracial (white-Negro), and intraindustry (skilled-unskilled) differences.

ciency. Still, the contribution of improvements in static efficiency to economic growth should not be overrated. In a study of an underdeveloped economy, Chile, Arnold Harberger suggested that the welfare cost of inefficient resource allocation may amount to 15 percent of national income of which approximately two-thirds would be due to inter-industry differences in the marginal productivities of the factors of production [10, p. 142]. Over a period of twenty years, the elimination of the remaining one-third would then raise a growth rate of 4.0 percent by less than one-third of one percentage point.

A further consideration is that while in measuring total factor productivity by the use of an aggregate production function we customarily assume constant returns to scale, increasing returns may be obtained at the plant, firm, industry, or national economy level. At the same time, with various forms of economies internal and external to the industry being exploited at different output levels, large-scale economies appear to be generally more important at earlier than at later stages of the process of industrialization. However, following Schumpeter, we may regard increasing returns to scale as dynamic phenomena—and subsume these under dynamic efficiency—if large-scale economies involve irreversible shifts in supply curves.

These considerations, as well as the possible relationship between investment and technological change, need to be taken into account in attempting to derive conclusions in regard to the dynamic efficiency of national economies and, in particular, the Soviet economy on the basis of estimates of total factor productivity.

#### IV

Estimates of total factor productivity in the Soviet Union have recently been made by Professor Bergson [3] whose results are reproduced in Table 1. The concept of output used in Bergson's calculations is net national product, and the inputs are labor, reproducible fixed capital, farm land, and livestock herds. National product has been valued at constant prices of the base year (1937) as well as in so-called "given-year" values.<sup>5</sup> Further, the services of reproducible fixed capital and livestock have been estimated, depending on the variant used, by using an interest rate of 8 and 20 percent, respectively.<sup>6</sup>

Bergson's results for the national economy of the Soviet Union in the period 1928-58 differ little from results obtained by Raymond Powell with respect to Soviet industry taken by itself. At base year ruble factor

<sup>5</sup> In the comparison of each year under consideration with the base year, output is valued at given-year prices.

<sup>6</sup> Bergson uses two variants with respect to land, too, but given the small differences shown in the results, this procedure has not been followed in the present paper.

costs, an annual rate of increase of total factor productivity of 1.2 and 1.3 percent is indicated in the two cases, respectively, if a 20 percent interest rate is used in the calculations, while a rate of increase of 1.7 percent is obtained in both instances if an 8 percent interest rate is assumed [3, p. 7] [14, p. 172].

The similarity in the results is puzzling since available information suggests that while improvements in production methods have taken place in Soviet industry in the period under consideration, improvements in agriculture, trade, and transportation have been rather limited. The results can be rendered more compatible, however, if Bergson's estimates are appropriately adjusted.

To begin with, adjustments need to be made for changes in the occupational structure of the labor force. In the Soviet Union, as in other industrializing countries with a large rural population, institutional rigidities as well as the existence of some degree of short-run complementarity between labor and capital in industry hindered the equalization of the marginal product of labor in individual sectors. Interindustry disparities in the quality of labor provide a further reason for differences in wages (marginal products).

Given the observed differences in the marginal productivity of labor between agriculture and industry in the Soviet Union, and the movement of labor from agriculture to industry in the period under consideration, the dynamic efficiency of the Soviet economy is overstated if an unweighted index of employment is used in the calculations. On the one hand, the movement of labor due to the removal—or lessening—of institutional rigidities leads to a once-for-all improvement which is properly classified as a change in static efficiency; on the other, the increased weight of skilled occupations represents an increase in inputs. At the same time, the alleged undervaluation of agricultural output in the Soviet Union cannot be invoked to justify the use of an unweighted employment index, since agricultural production has been “undervalued” in calculating the output index, too.

More generally, one may object to the application of different procedures in weighting outputs and inputs: if outputs are weighted by ruble factor costs (the sum of direct and indirect labor costs), inputs should be weighted accordingly. With given product prices, the appropriate procedure involves weighting employment in the various sectors by wages paid in the base year.<sup>7</sup> In the present paper, we have adjusted for the shift of labor from agricultural to nonagricultural occupations where the

<sup>7</sup> The proposed adjustment entails an upward correction in the input index, leaving the output index unchanged. On the other hand, if intersectoral differences in marginal productivities were due solely to the undervaluation of agricultural output, the appropriate adjustment would involve a downward correction in the output index without changing the input index. Here again measured increases in total factor productivity would be reduced.

movement of labor was of greatest importance.<sup>8</sup> The relevant employment indices and the derivation of wage shares in the base year are given in the Appendix tables while the results are shown, alongside with those of Professor Bergson, in Table 1.

The separation of agricultural and nonagricultural employment modifies the estimated values of total factor productivity for the Soviet economy to a considerable degree, with much of the adjustment pertaining to the early part of the period. Professor Bergson indicated that the inclusion of inventories also affects the results to some extent [3, p. 22]. Conceptually, the services of inventories should be considered a capital input and are regarded as such by Raymond Powell. International comparability, too, requires the inclusion of inventories in calculating changes in total factor productivity in the Soviet economy, since the services of inventories are comprised in the profit figures used to represent the services of capital in free enterprise economies.<sup>9</sup>

After appropriate adjustments, total factor productivity in the Soviet economy appears to be rising at a rate lower than in industry taken by itself. Using 1937 ruble factor costs as weights, estimates for the period 1928-58 indicate an annual rate of increase in total factor productivity of 0.5 or 0.9 percent for the Soviet economy as a whole and 1.3 and 1.7 percent for industry, depending on whether a 20 percent or an 8 percent interest rate is used in valuing the services of capital.

Although some differences are observed in the procedures used by Bergson and Powell, the methods employed are sufficiently similar to enable us to conclude that there has been a significant difference between the performance of the Soviet economy taken as a whole and that of Soviet industry.<sup>10</sup> A comparison with estimates made by D. Gale Johnson for Soviet agriculture strengthens this conclusion. With 1955 as the base year, and employing an interest rate of 8 percent in valuing the services of capital, Johnson's results show an annual rate of increase of total factor productivity of 0.2 or 0.8 percent in Soviet agriculture during the period 1928-58, depending on whether labor is measured in terms of man-days or numbers employed [11, p. 218].<sup>11</sup>

<sup>8</sup> On the other hand, no adjustment has been made for the United States where the inter-industry shift in employment appears to have little importance. Concerning this point, see [12, p. 34].

<sup>9</sup> This observation does not apply to U.S.-U.S.S.R. comparisons made by Bergson, since he excluded the services of inventories in regard to the United States, too. But whereas the inclusion of inventories in the aggregate production function does not modify the results obtained for the United States, the estimates for the Soviet Union are affected thereby.

<sup>10</sup> Problems related to the estimation of gross national product and industrial production hinder the comparability of data relating to subperiods ending and starting in 1940, however.

<sup>11</sup> The corresponding results are 0.8 and 1.4 percent if 1928 weights are used in regard to current purchases that increased from 1.6 percent of agriculture income in 1928 to 8.2 percent in 1938 and declined again to 5.6 percent in 1955 [11, p. 219]. But the use of 1928 weights in calculating a combined index of inputs would raise the estimated value of total factor productivity on the national economy level, too.



## V

We come next to a comparison of the performance of the Soviet and the American economies. As the estimates of Table 1 indicate, for the period taken as a whole, the Soviet performance appears to be inferior to that of the U.S. economy if Soviet output is valued in 1937 ruble factor costs, while the opposite conclusion is reached if, for the U.S.S.R., given-year values are used as weights.

It appears, then, that the results depend to a considerable extent on the choice of factor-cost weights. In a pathbreaking article, Richard H. Moorsteen has indicated the superiority of given-year weights for cases when the number of observations exceeds two [13]. In Professor Bergson's calculation given-year weights were used in regard to 1928 and 1950 data, with 1937 as the base year. In the comparison between 1937

TABLE 1

U.S.S.R. AND U.S.A.: NET NATIONAL PRODUCT, FACTOR INPUTS AND PRODUCTIVITY,  
AVERAGE ANNUAL RATES OF GROWTH FOR SELECTED PERIODS  
(Percent)

	1928-58	1928-40	1940-50	1950-58
<i>Soviet Union</i>				
Net national product,				
in 1937 ruble factor cost.....	4.1	4.2	1.9	6.8
as composite 1937 base.....	6.0	9.3	1.7	6.8
Total factor productivity I				
with output in 1937				
ruble factor cost A.....	1.2	0.1	1.3	2.7
B.....	1.7	0.5	1.2	4.0
with output in given				
year values A.....	3.0	4.9	1.1	2.7
B.....	3.5	5.3	1.0	4.1
Total factor productivity II				
with output in 1937				
ruble factor cost A.....	0.6	-1.0	0.9	2.7
B.....	1.0	-0.8	0.8	3.9
with output in given				
year values A.....	2.5	3.8	0.7	2.7
B.....	2.9	4.1	0.6	3.9
Total factor productivity III				
with output in 1937				
ruble factor cost A.....	0.5	-1.1	0.8	2.5
B.....	0.9	-0.9	0.8	3.7
with output in given				
year values A.....	2.4	3.7	0.6	2.5
B.....	2.8	4.0	0.6	3.7
	1869/78- 1899/1908	1899/1908 -1929	1929-57	1948-57
<i>United States</i>				
Net national product.....	4.6	3.4	2.9	3.7
Total factor productivity.....	1.5	1.8	2.0	1.7

SOURCES: *ETSU*, pp. 6-7 and Appendix Table 4.

NOTES: Variant A: 20 percent annual net return assumed for reproducible fixed capital, inventories, and livestock.

Variant B: 8 percent annual net return assumed for reproducible fixed capital, inventories, and livestock.

I. Unadjusted.

II. Adjusted for intersectoral shifts in employment.

III. Adjusted for intersectoral shifts in employment and changes in inventories.

and 1950, the choice of weights hardly affects the results; hence we are left with a binary comparison between the years 1928 and 1937.

According to the estimates of Appendix Table 4, total factor productivity in the Soviet economy declined between 1928 and 1937 if output is valued in 1937 (base-year) ruble factor costs while an increase is shown if 1928 (given-year) factor costs are used as weights. Thus, if we adjusted for changes in inputs by appropriately "scaling down" the transformation curve of the year 1937, it would appear that the production possibilities of the Soviet economy in 1937 did not include the output-mix of the year 1928 and vice versa. These results are explained, in part, by the adverse effects of the policy of collectivization on agricultural productivity that were accompanied by an increase in the relative price of food and a decline in the share of agriculture products in the output-mix.<sup>12</sup>

A divergence between the results calculated in 1928 and in 1937 prices is also observed if we restrict our investigation to Soviet industry. Using an interest rate of 20 percent in valuing the services of capital, Raymond Powell estimated that, between 1928 and 1940, total factor productivity in Soviet industry increased at an annual rate of 1.5 percent if output is valued in 1937 prices and 6.6 percent in 1928 price [14, p. 172]. Part of this discrepancy is again explained by the poor performance of agriculture which led to high relative prices and low consumption of food products. Another important influence appears to be the decline of input-coefficients in the machinery industry. The latter changes can be largely accounted for by reference to large-scale economies that could be appropriated in the process of transformation of Soviet industry from small-scale establishments into large units and the borrowing of Western technology.

The exploitation of large-scale economies and the ready availability of foreign technology provided an advantage to the Soviet economy as against the U.S. economy of the same period. Hence, comparison should be properly made with the corresponding stage of U.S. industrial development. Professor Bergson suggested that the period 1869-78-1899-1908 in the United States may be suitable for this purpose. Using 1929 prices, Bergson estimated total factor productivity in the United States to rise at an annual rate of 1.5 percent during this period. Adjusting for the movement of labor from agricultural to nonagricultural occupations, the relevant figure is 1.3 percent. By comparison, the estimated rate of change of total factor productivity is -0.9 to -1.1 percent in the

<sup>12</sup> According to D. Gale Johnson's estimates, agricultural output rose by 4 percent between 1928 and 1938 while, calculating with 1955 weights, inputs appear to have increased by 41 or 26 percent, depending on whether labor is measured in terms of man-days or numbers engaged. (The increase of inputs was 19 and 7 percent, respectively, if 1928 weights are used in regard to current purchases.) Output reached 105 percent of the 1928 level in 1940 but data on inputs are not available for that year [11, pp. 210, 218].

period 1928-40 in the Soviet Union if output is valued in 1937 rubel factor costs and 3.7 to 4.0 percent if 1928 factor costs are used as weights (Table 1).

But the latter estimate can hardly serve for purposes of comparison since its counterpart in the United States would have to be weighted at 1869-78 prices. In fact, in view of Gerschenkron's law, the base-year weighted measure would also appear to understate increases in total factor productivity during the period of industrialization in the United States by reason of the use of 1929 prices in the calculations.

Furthermore, during the 1930's the Soviet Union had the advantage of being able to draw on an accumulated fund of technical knowledge that was not available to the United States in the nineteenth century. It appears, therefore, that, in terms of base-year prices, the performance of the United States was superior to that of the Soviet Union at the same stage of industrialization. Available information does not permit reaching a conclusion with regard to changes measured in given-year prices, however.<sup>13</sup>

On the other hand, the choice of price weights hardly affects the results obtained for the 1950's and neither does the shift of labor from agriculture to industry. Soviet data referring to these years can be compared to calculations made for several capitalist countries in a study conducted at M.I.T. Especial interest attaches to the comparison of the performance of the Soviet and the Japanese economies, since these countries are at similar levels of economic development and both could rely on borrowing advanced technology from abroad.

In order to conform to the procedure used in the M.I.T. study, no adjustment was made for the intersectoral shift of labor in estimating changes in total factor productivity in the Soviet Union. Further, output has been measured as gross rather than net national product. Correspondingly, in deriving an index of combined inputs, capital has been weighted by its share in GNP; i.e., with the inclusion of an allowance for depreciation.

According to the results shown in Table 2, Japan, the Soviet Union, and Germany experienced the highest rate of increase of total factor productivity in industry, with the United States, Canada, and the United Kingdom following. Germany and Japan appear to be in the lead in regard to the growth of total factor productivity in the national economy as a whole, the Soviet Union falls somewhat behind, while the ranking of the other three countries does not change.

<sup>13</sup> On the other hand, Professor Bergson argues that "the shift of the base year for the United States even over a long interval to a relatively early date might not have as large an impact on real national income as that observed for the U.S.S.R. when the base year is shifted from 1937 to 1928" [3, p. 15]. But no evidence is offered in support of this proposition; hence—as Bergson also notes—there appears to be need for further inquiry into this matter.

TABLE 2

SELECTED COUNTRIES: NET NATIONAL PRODUCT, FACTOR INPUTS AND PRODUCTIVITY,  
AVERAGE ANNUAL RATES OF GROWTH FOR SELECTED PERIODS

(Percent)

	Soviet Union		United States	Canada	United Kingdom	Germany	Japan
	1950-58		1948-57	1949-60	1949-59	1950-59	1951-59
	A	B					
GNP at factor cost.....	7.2	7.2	3.8	4.0	2.4	7.2	8.4
Selected inputs.....	4.8	3.7	2.3	2.8	1.8	3.6	4.6
Total factor productivity...	2.3	3.5	1.5	1.2	0.6	3.6	3.7
Industrial production (gross value added).....	10.1	10.1	4.0*†	3.7*	3.3*	8.8	11.6*
Selected inputs.....	6.7	6.2	1.4	2.3	2.6	5.3	7.4
Total factor productivity...	3.3	3.8	2.6	1.4	0.7	3.4	4.1

\* Manufacturing output. † 1949-60.

SOURCES: GNP—U.S.S.R. and U.S.—Table 1, with adjustment made for depreciation as given in *ETSU*, pp. 13, 36.

Other countries—Domar *et al.*, "Economic Growth and Productivity in the United States, Canada, United Kingdom, Germany and Japan in the Post-War Period," *Rev. of Econ. and Statis.*, Feb., 1964.

Industrial production—

U.S.S.R.—R. P. Powell, "Industrial Production," *Economic Trends in the Soviet Union*, pp. 155, 171-72.

Other countries—Domar *et al.*, *op. cit.*

Several caveats are necessary in interpreting these estimates, however. To begin with, the period under consideration is relatively short and, hence, errors in observation can have a substantial effect on the comparison of terminal years. Also, the assumption of unembodied technological change may be less realistic for a short time period than over a period of several decades, especially if an acceleration in the growth of capital changes the age composition of the capital stock to a considerable extent and the new vintage of capital embodies advanced techniques.

These observations appear to have relevance for the three fastest-growing economies, Japan, Germany, and the Soviet Union, where an acceleration of the growth of capital stock has been experienced during the 1950's. In the U.S.S.R., the capital stock grew at an average annual rate of 11.2 percent, in Japan 8.8 percent, and in Germany 6.8 percent, while in Canada the increase was 5.5 percent, in the United States 3.7 percent, and in the United Kingdom 3.5 percent a year. At the same time, in the three countries that showed the best performance, much of equipment was antiquated by reason of the lack of replacement during World War II, and hence the new capital goods were of considerably higher quality than the old installations.

The results obtained with respect to the Soviet economy have also

been affected by the exceptionally good harvest in 1958. On the basis of available information on recent developments in agriculture as well as in industry, it may be suggested that the measured increase of total factor productivity would be lower if a later terminal year were used in the calculations. This conclusion is strengthened if account is taken of productivity gains made in postwar reconstruction and in releasing workers from forced labor camps during the 1950's.<sup>14</sup>

The above considerations indicate the difficulties encountered in interpreting the results of the present calculations. It is hoped, however, that the increased availability of data as well as improvements in statistical methodology will increase the reliability of estimation in the future. Until then, the conclusions given here should be considered as highly tentative.

As regards the factors contributing to economic growth, our results appear to suggest that increases in inputs and improvements in static efficiency have been of considerable importance in the Soviet Union. With the capital stock increasing about sevenfold and output reaching 336 percent of the 1928, capital formation should be accorded first place and the increase of industrial employment associated with the shift of labor from agricultural to nonagricultural occupations second. Improvements in dynamic efficiency appear to have assumed importance after 1950, although the exploitation of economies of scale and the borrowing of foreign techniques contributed to a reduction of input coefficients already during the 1930's.

However, in terms of dynamic efficiency, the performance of the Soviet economy does not appear to be superior to that of free enterprise economies at a similar level of economic development. Rather, the achievements of the Soviet economy should be explained by its success in attaining a high rate of capital formation and in reallocating labor from agriculture to industry.

#### REFERENCES

1. Kenneth Arrow, H. B. Chenery, B. Minhas, and R. M. Solow, "Capital-Labor Substitution and Economic Efficiency," *Rev. of Econ. and Statis.*, Aug., 1961, pp. 225-50.
2. Bela Balassa, *The Hungarian Experience in Economic Planning* (Yale Univ. Press, 1959).
3. Abram Bergson, "National Income," *Economic Trends in the Soviet Union*, A. Bergson and S. Kuznets, eds. (Harvard Univ. Press, 1963), pp. 1-37. Also cited as *ETSU*.
4. ———, *The Real National Income of Soviet Russia Since 1928* (Harvard Univ. Press, 1961). Also cited as *Real SNIP*.
5. ———, "Socialist Economics" in *A Survey of Contemporary Economics*, Vol. 1, H. S. Ellis, ed. (Irwin, 1948), pp. 412-48.
6. Abram Bergson, Hans Heymann, Jr., and Oleg Hoeffding, *Soviet National Income and Product, 1928-1948: Revised Data*, The RAND Corp. Research Memorandum, RM-2544, Nov. 15, 1960, p. 8. Also cited as *Real SNIP, 1928-1948*.
7. Evsey D. Domar, *et al.*, "Economic Growth and Productivity in the United States,

<sup>14</sup> The former of these factors has been of importance also in Germany and Japan.

- Canada, United Kingdom, Germany and Japan in the Postwar Period," *Rev. of Econ. and Statis.*, Feb., 1964.
8. Evsey D. Domar, "On Total Productivity and All That," *J.P.E.*, Dec., 1962, pp. 597-608.
  9. R. Dorfman, P. A. Samuelson, and R. M. Solow, *Linear Programming and Economic Analysis* (McGraw-Hill, 1958).
  10. Arnold C. Harberger, "Using the Resources at Hand More Effectively," *A.E.R.*, May 1959, pp. 134-46.
  11. D. Gale Johnson, "Agricultural Production," *Economic Trends in the Soviet Union*, pp. 203-34.
  12. J. W. Kendrick, *Productivity Trends in the United States* (N.B.E.R., Princeton Univ. Press, 1961).
  13. R. H. Moorsteen, "On Measuring Productivity Potential and Relative Efficiency," *Q.J.E.*, Aug., 1961, p. 451.
  14. Raymond P. Powell, "Industrial Production," *Economic Trends in the Soviet Union*, pp. 150-202.
  15. R. M. Solow, "Investment and Technical Progress," *Mathematical Methods in the Social Sciences*, K. J. Arrow, S. Karlin, and P. Suppes, eds. (Stanford Univ. Press, 1960), pp. 89-104.
  16. ———, "Technical Change and the Aggregate Production Function," *Rev. of Econ. and Statis.*, Aug., 1957, pp. 312-20.

## APPENDIX

TABLE 1

U.S.S.R.: ESTIMATES OF EMPLOYMENT

	1928	1937	1940	1950	1958
Employment, millions*	49.4	68.7	81.5	84.1	93.6
Farm.....	35.0	37.4	41.5	38.5	37.9
Nonfarm.....	14.4	31.3	40.0	45.6	55.7
Working hours, nonfarm, index†	103	100	109	118	113
Employment, millions					
Adjusted for working hours.....	49.8	68.7	85.1	92.3	100.8
Farm.....	35.0	37.4	41.5	38.5	37.8
Nonfarm.....	14.8	31.3	43.6	53.8	62.9
Employment, millions					
Adjusted for working hours‡ and boundaries.....	49.8	68.7	76.8	82.7	91.0
Farm.....	35.0	37.4	37.4	34.5	34.2
Nonfarm.....	14.8	31.3	39.4	48.2	56.8

\* *Real SNIP*, p. 443; *ETSU*, p. 16.† *Real SNIP*, p. 425.‡ *ETSU*, p. 4.

TABLE 2

U.S.S.R.: WAGES AND LABOR INCOME, 1937

	Employment* (millions)	Per Capita Income (rubles)	Wage and Labor Income (billion rubles)
Collective farmers.....	32.8	1607	52.7†
State farms.....	2.0	2121‡	4.2
MTS.....	0.4	2121‡	0.8
Hired labor.....	0.3	1607§	0.5
Independent peasants.....	1.9	2315‡	4.4†
Farm Employment and Incomes.....	37.4	1674	62.6
Adjustment for social insurance.....	—	—	0.3**
Farm employment and income including social insurance.....	37.4	1682	62.9
Nonfarm employment and labor incomes.	31.3	3738	117.0††
Total employment and labor incomes..	68.7	2619	179.9

\* *Real SNIP*, Appendix K, pp. 443, 446.† *Ibid.*, p. 118.‡ *Ibid.*, p. 121.

§ Assumed to be equal to average labor income in collective farms.

\*\* Calculated on the basis of the reported 6.6 billion rubles of employers' social security contributions (*SNIP 1928-48*, p. 8) under the assumption that social insurance contributions are proportional to the wage bill in the state sector.†† *ETSU*, p. 19.

TABLE 3  
U.S.S.R.: WEIGHTS USED IN AGGREGATING FACTOR INPUTS

	WEIGHTS A		WEIGHTS B	
	Billion Rubles	Percent	Billion Rubles	Percent
Employment.....	179.9	70.4	179.9	80.4
Farm.....	62.9	24.5	62.9	28.0
Nonfarm.....	117.0	45.9	117.0	52.4
Reproducible fixed capital....	46.2	18.1	18.5	8.3
Farm land.....	22.4	8.8	22.4	10.0
Livestock herds.....	7.0	2.7	2.8	1.3
All inputs.....	255.5	100.0	223.6	100.0

SOURCES: *ETSU*, p. 19 and Appendix Tables 1 and 2.

NOTES: Variant A: 20 percent annual net return assumed for reproducible fixed capital, inventories, and livestock.

Variant B: 8 percent annual net return assumed for reproducible fixed capital, inventories, and livestock.

TABLE 4  
U.S.S.R.: NATIONAL PRODUCT, FACTOR INPUTS, AND FACTOR PRODUCTIVITY, SELECTED YEARS  
(1937=100)

		1928	1937	1940	1950	1958
Net national product	In 1937 ruble factor cost.....	67.1	100.0	110.3	132.7	225.5
	In given year values.....	38.1	100.0	110.3	129.9	221.0
Employment, all sectors	Number of workers adjusted for changes in nonfarm hours..	72.5	100.0	111.8	120.4	132.5
Employment, farm	Number of workers.....	93.6	100.0	100.0	92.3	91.4
nonfarm	Number of workers adjusted for changes in nonfarm hours..	62.4	100.0	130.2	154.0	181.5
Reproducible fixed capital	In 1937 rubles.....	41.0	100.0	126.4	133.2	311.4
Farm land	Acres.....	83.5	100.0	101.1	98.3	131.4
Livestock herds	In 1937 rubles.....	147.3	100.0	117.5	108.0	139.4
Selected inputs, unadjusted	With 1937 weights A.....	69.7	100.0	113.7	120.4	165.1
	B.....	71.9	100.0	111.9	119.1	147.4
Selected inputs, adjusted for intersectoral shifts in employment	With 1937 weights A.....	63.3	100.0	117.3	129.0	177.5
	B.....	64.7	100.0	116.1	128.8	161.4
Total factor productivity, unadjusted	With output in 1937 A.....	96.3	100.0	97.0	110.2	136.6
	Ruble factor cost B.....	93.3	100.0	98.6	111.4	153.0
	With output in given year A..	54.7	100.0	97.0	107.9	133.9
	Values B..	53.0	100.0	98.6	109.1	149.9
Total factor productivity, adjusted for intersectoral shifts in employment	With output in 1937 A.....	106.0	100.0	94.0	103.0	127.0
	Ruble factor cost B.....	103.7	100.0	95.0	103.2	139.7
	With output in given year A..	60.2	100.0	94.0	100.7	124.5
	Values B..	58.9	100.0	95.0	100.9	136.9

SOURCE: *ETSU*, pp. 4-5 and Appendix Tables 1-3.



## INTERNATIONAL DIVISION OF LABOR IN CEMA: LIMITED REGRET STRATEGY

By EGON NEUBERGER\*  
*The RAND Corporation*

### I. *Introduction*

The members of the Council for Economic Mutual Assistance (CEMA) are very far from achieving an optimum division of labor if their performance is judged by the yardstick of traditional Western economic theory or by the degree to which it implements the objectives of CEMA leaders. The thesis of this paper is that the "limited regret strategy," a conservative strategy that describes CEMA efforts, may nevertheless be considered a reasonable strategy when full account is taken of conflicting objectives, political, economic, and ideological constraints, and a lack of satisfactory theory of division of labor in the East or the West. In view of these serious impediments to the achievement of a more effective international division of labor, the relatively small steps taken thus far in CEMA appear, on the whole, to be sensible.

### II. *Conflicting Objectives and Constraints*

Every attempt to achieve a complete division of labor faces numerous obstacles; in the case of CEMA these obstacles are probably more inhibiting than in most other cases. The three major types of obstacles are: the desire to accomplish a multiplicity of objectives; the inability or unwillingness to overcome certain political, economic, and ideological constraints; and the absence of a satisfactory theory of division of labor.

*Objectives.* The main motives behind Soviet efforts to achieve a division of labor in CEMA probably include the desire for greater political cohesion (partially through the means of the creation of greater economic interdependence among CEMA countries) and the initiation of a movement toward the goal of a unified Communist commonwealth. The primary motive of the other CEMA members is probably the possibility of economic gains. There are many different avenues of achieving economic gains from a division of labor. Examples are: a freer flow of factors of production, ideas, and technologies; effects of an enlarged market on competition among productive units; a possible reduction in costs based on an increase in the size of these units; a concentration on commodities that can be produced most efficiently; and so on. In

\* This research is sponsored by the United States Air Force under Project RAND.

this study we shall concentrate on the last two aspects of the efficient use of resources: the specialization on products in which a country has a comparative advantage and the reaping of economies of scale.

The objectives that impede the movement toward a more comprehensive division of labor in the CEMA countries are listed below.

In addition to the twin goals in CEMA of achieving the maximum benefit for the group as a whole and for each individual country, there is even a third goal which has been formally accepted by all member governments: the closing of the gap in the levels of development between the economically more advanced and less advanced members. The stated purpose is to enable all the countries to enter into the stage of full communism within approximately the same historical period. The first two goals are not necessarily complementary, and the third one may conflict with the first and will certainly conflict with the second. Thus, what is needed, to use Marxist terminology, is a true dialectical approach to the problem.

CEMA leaders, as well as leaders of many underdeveloped countries, accept the necessity of developing a highly diversified national economy in their own country; that is, the development of all major sectors of the economy in each country. They oppose any division of labor that might be regarded as leading to overspecialization.

Closely connected with this is the dogma of the need for priority development of producers' goods industries in each country. This represents a reinterpretation of Marx's theory of the need for more rapid development of producers' goods industries due to insufficiencies of consumer demands in capitalism.

CEMA leaders place even greater stress than is usual on the desirability of maintaining full employment of labor and plant at all times.

*Constraints.* There are two main sets of constraints inhibiting the movement toward a division of labor. One set is based on the existence of separate countries, each with its own jealously guarded sovereignty, and the other set arises out of the existence of Soviet-type planned economies in each of the countries.

As we argued in a previous study,<sup>1</sup> the sovereignty issue may be more sensitive in CEMA than in other cases because of the relative power of the U.S.S.R. compared with other members, the past history of Soviet relations with them, the question of the legitimacy of satellite governments, and the greater importance of control over the economy as a basis of power in centrally planned economies.

There are four major constraints connected with the sovereignty issue: (1) a division of labor is feasible only when it provides a clear

<sup>1</sup> Egon Neuberger, *Soviet Bloc Economic Integration: Some Suggested Explanations for Slow Progress* (The RAND Corp., RM-3629-PR, July, 1963).

benefit to each and every one of the participating countries; (2) no division of labor is possible which results in a serious imbalance in the current account of the balance of payments of any country;<sup>2</sup> (3) the mobility of resources across national boundaries is even more limited in CEMA than in other groups of countries; and (4) CEMA countries find it difficult to adopt a unified conception of division of labor for CEMA as a whole. A Czech economist complained to the author that no real progress was possible until this was done, but that it cannot be accomplished because it is regarded as a first step toward a unified planning mechanism which some political leaders and some planning officials abhor.<sup>3</sup>

There are also four main constraints arising out of the centrally planned nature of CEMA economies:

1. The planners must take specific, active measures to implement a division of labor instead of being able to concentrate on removing artificial barriers, such as tariffs, and then permitting individual firms to act upon price and cost differentials. As argued previously,<sup>1</sup> this task places a greater burden of detailed decision making and bargaining on the planners and makes the achievement of a division of labor much more difficult and time consuming. This is not to say that the market will necessarily yield better results in the welfare sense but rather that planners must ask themselves questions that private entrepreneurs need not ask.

2. In Soviet-type planned economies, domestic prices in each country do not play a primary role in allocating resources and do not reflect real opportunity costs as viewed by the leaders. This is true even within the context of the labor theory of value and much more so within the context of Western economic theory. There are differences among the countries in the methods of calculating prices and costs, and the official exchange rates are set quite arbitrarily. Thus, prices and costs in each country, brought to a common denominator by using existing exchange rates, cannot be used for guiding planners' decisions; this fact is well recognized in CEMA. On the other hand, CEMA leaders are not willing to discard their planning system and substitute a system based on the market mechanism. Because of the adherence to the Marxist labor theory of value, even if fundamental price reforms were undertaken in each country, the prices would probably not be useful tools to planners in their attempts to develop a rational (in terms of the objectives functions of the leaders) division of labor among the countries. Even these

<sup>2</sup> The first two constraints would be considerably less inhibiting if a system of large inter-country compensation payments were feasible; the author does not believe that it is.

<sup>3</sup> Interview with Jiri Novozamsky, Chief of the Section on the World Socialist System, Scientific Research Institute on Planning of the State Planning Commission, Prague, Sept. 5, 1963.

new prices would continue to undervalue the contribution made by the very scarce capital and land resources in each country.

3. The existence of planned systems and the irrationality of price systems lead to a strong preference for bilateral trade relations instead of a system of multilateral trade with freely convertible currencies. Planners feel they can compensate partially for price irrationality by bartering market baskets of goods where the two market baskets approach more closely to equality in terms of the leaders' utilities. This would not be possible if goods were bought and sold freely for convertible currencies at irrational domestic prices, at prices reigning in intrabloc trade, or at world market prices.

4. Vested interests always present obstacles to the achievement of a division of labor. In the case of CEMA the vested interests are, in fact, government economic institutions. It is possible that they may be in a position to exert greater pressure against any change in the status quo than firms in free enterprise economies.

The obstacles to a division of labor that are more serious in the case of CEMA than in other cases are: the inconsistency of the objectives; the lack of resource mobility; the unwillingness to use the market mechanism while prices and exchange rates are inappropriate for use even as planning tools; the stress on development of almost every sector of the economy in each country; the priority given to producers' goods industries; the use of the labor theory of value; and the strong preference for bilateral trade relations.

### III. *Theory of International Division of Labor*

None of the theories of international division of labor from Smith and Ricardo to Ohlin, Haberler, and Viner provide a completely satisfactory answer to the problem even in the context within which they were developed.<sup>4</sup> The tools that they provide can certainly not be applied mechanically in the world of the Soviet bloc as defined by the constraints just listed. Under present conditions of price formation in CEMA it would not be possible to apply the doctrine of comparative advantage, even if the desire existed, which it does not. The schizophrenic attitude toward this doctrine in the Soviet bloc is discussed in Professor Pryor's excellent article.<sup>5</sup> Prices in CEMA countries fail to

<sup>4</sup> Considerable empirical and theoretical work in this field has always been going on. Two very recent examples are: Bela Balassa, "An Empirical Demonstration of Classical Comparative Cost Doctrine," *Rev. of Econ. and Statis.*, Aug., 1963, pp. 231-38, an empirical study of the classical hypothesis that division of labor is based on productivity differentials; and J. L. Ford, "The Ohlin-Heckscher Theory of the Basis of Commodity Trade," *Econ. J.*, Sept., 1963, pp. 458-76, a discussion of the limitations inherent in the Ohlin doctrine that differences in factor endowments are the basis of a division of labor.

<sup>5</sup> Frederic L. Pryor, "Foreign Trade Theory in the Communist Bloc," *Soviet Studies*, July, 1962, pp. 41-61. Many other points in this study are dealt with in his book, *The Communist Foreign Trade System* (Massachusetts Inst. of Tech. Press, 1963).

include charges for the services of capital and land, underestimate depreciation and depletion; include completely arbitrary profit and turnover tax rates, and do not always value different qualities of labor by their opportunity costs.

The theories developed by Communist economists fit more closely the institutional framework within the bloc, but do not even approach the stage of representing a satisfactory solution of the problem.<sup>8</sup> The bias in CEMA is toward the use of absolute cost differentials among countries as the primary measure of the appropriate type of specialization decisions. But for this purpose, in addition to the problem of prices and costs that do not reflect opportunity costs in each country, there are problems of differences in costing methods among the countries and the arbitrary nature of exchange rates. If these institutional obstacles could be overcome, it would be possible to approximate the results of the comparative cost approach by a ranking of countries for each commodity according to their absolute costs, expressed in a common unit of account; for example, the ruble. The specialization assignment in each commodity would be given to the country with the lowest costs, and then exchange rates altered until each country obtained assignments in a sufficient number of commodities to assure it a balance in its international accounts. However, neither the comparative cost nor this "absolute cost" approach solves the problems raised by the lack of appropriate theoretical tools to handle the need to take into account future costs rather than present costs, rapid changes in technology, capital endowment and labor skills, the need to consider the changes in costs due to internal economies of scale based on specialization, the problem of taking into account the external economies and diseconomies and the interdependencies among different industries, and the fact that even the final bill of goods desired by the leaders is likely to change over time, in part due to different income elasticities of demand for various consumer goods and even for government expenditures.

An aspect that has not been dealt with satisfactorily in bloc theory is the effect on the international socialist division of labor of the existing alternative of trading with nonmember countries. The implications of this issue are too broad to be dealt with summarily in this study.

#### IV. *Limited Regret Strategy*

*Choice of Strategy.* In a particular case when planners must determine whether to adopt a policy that would result in a greater division of labor, one appropriate strategy would be to adopt this policy if the expected gains are greater than the expected losses. Given several policies, the one yielding the largest net gain should be selected.

<sup>8</sup> The basic document is the "Basic Principles of the International Socialist Division of Labor," *Pravda*, June 17, 1962, pp. 3-4.

Such a strategy, however, is not appropriate to the case of CEMA, and a different approach seems to be followed by CEMA leaders in altering national plans. These plans have traditionally been the bases for managing the economy, and any changes in them must yield expected benefits significantly greater than expected losses in order to be considered acceptable. The extreme stress on national sovereignty and the development of a diversified economy in each country make the political and economic risks of excessive specialization loom large in the minds of both the political leaders and the planners. It is likely that a planner who makes a mistake in providing for too much autarky is less likely to suffer in terms of power or prestige than a planner who errs in the other direction. The inability to measure gains and losses with any accuracy, the lack of an adequate theory of division of labor, and the danger of placing excessive reliance on other planned economies for timely deliveries of high quality goods—all create further biases against extensive specialization. The lack of flexibility of adjustment to unforeseen circumstances in centrally planned economies raises the specter that extensive specialization could lead to serious disruptions in national plans. In view of all these factors, it is probable that CEMA decision-makers will prefer a relatively conservative strategy, motivated by risk aversion. We shall call this the strategy of limited regret.

In this strategy, a policy that involves an adjustment in national plans will be adopted only if it meets two conditions. The first condition is that the expected gains from the policy exceed the expected losses by some minimum positive amount sufficient to justify a change from the *status quo*, where the weights attached by the decision-makers to the expected values of the undesirable outcomes are higher than those attached to desirable outcomes. The second condition is that the results of the worst possible set of expected outcomes are not considered as too damaging, or if they are, that the probability of these outcomes occurring is sufficiently low to make the risk acceptable.

*CEMA Actions.* What is the actual approach by which a division of labor is brought about in CEMA? The approach consists of three major steps.

The first step is the preparation of long-term national plans by each country based primarily on its own resources and on the bill of goods desired by its own national leaders. After a first draft of each plan is completed, the countries enter into two types of negotiations in order to obtain gains from an international division of labor. These negotiations result in marginal adjustments to the national plans. Step two consists of a series of bilateral discussions on the possibilities of mutually advantageous imports and exports during the plan period. Step three, generally taking place concurrently with the second step, involves multilateral discussions on specialization in each of the specialized

standing committees of CEMA. Each committee deals with one branch of the economy, such as the coal industry, the chemical industry, construction, agriculture, and so on<sup>7</sup>, and each has subcommittees dealing with a given range of products or services within the branch. The committees and their subcommittees are the crucial loci of specialization decisions, with the discussions regarding the *quid pro quo* each country obtains being held mainly in the subcommittees, to a lesser extent in the committees, and to an even smaller extent at the level of the Executive Committee of CEMA. This institutional set-up is most suitable for implementation of the limited regret strategy because it lends itself most easily to reaching agreements on adjustments to plans based on comparisons of technological coefficients within given sectors of the economy without the need for value comparisons.

*Expected Benefits and Risks of Losses.* The major benefits of engaging in a division of labor are the possibility of greater political cohesion (although political difficulties may also arise) and a more rapid rate of growth of GNP. The latter would be due, in good part, to the improved use of resources in each country by specializing in commodities in which the country has a comparative advantage and from reaping economies of scale resulting from specialization.

In addition to the political risk of losing some sovereignty and the economic risk of possible disruption of national plans, there are three other types of economic risks, each of them involving a certain type of opportunity cost. The first type of risk consists of agreeing to a specialization assignment in a commodity in which the country does not have a present comparative advantage or where diseconomies of scale outweigh economies. The second type involves specializing in a commodity in which it now has a comparative advantage but will not have this advantage in the future. The third type consists of specializing in a commodity whose production involves important external diseconomies or where the production of some other commodity would have led to a more rapid rate of growth through external economies of scale, interdependencies, creation of new labor skills, greater degree of technological innovation, and so on.

The first type of risk is generally minimized in market economies because firms operate primarily on the basis of existing price relationships. The probability of avoidance of the second type of risk is much lower, while there is almost no consideration of the third type. The last two types of risks may be impossible to measure satisfactorily even in those economies where prices represent existing opportunity costs, and none

<sup>7</sup> For the list of committees and a brief discussion of their work, see A. Aleksandrov, "V postoiannykh komissiiakh Soveta Ekonomicheskoi Vzaïmopomoshchi" (In the Standing Committees of the Council for Economic Mutual Assistance), *Vneshniaia Torgovlia*, 6, 1963, pp. 15-20.

of the three are amenable to accurate measurement when prices do not reflect opportunity costs. In this latter case, the limited regret strategy would seem to be especially appropriate as compared with less conservative strategies. It is possible to suggest some rules of thumb that would tend to result in specialization assignments in those cases where none of the probable risks of losses are likely to be very high and the gains from specialization are likely to be large relative to the losses.

*Rules of Thumb.* There appear to be three rules of thumb meeting the above requirement:

1. Take the case of raw materials or industries in which a large proportion of the cost is based on weight-losing raw materials. If inputs of both labor hours and machine hours (of roughly equivalent qualities) per unit of output of commodity  $x$  are much lower in Country I than in Country II, it might be possible to determine that Country I should specialize in commodity  $x$  and export it to Country II; thereby gaining considerable benefits without running very serious risks of the first type. Thus, unless large risks of the second or third types appear likely, such a specialization agreement would be feasible. It should be made clear, however, that this comparison for only one commodity in isolation does not assure us that the more efficient country necessarily has a comparative advantage in this commodity; a comparison of relative costs of producing all commodities in Countries I and II may show that Country I actually has a comparative disadvantage in the production of  $x$ . However, if the physical input indexes show very large differences, the probability is high that Country I has a comparative advantage in  $x$  and this could serve as a useful rule of thumb in cases where relative prices do not represent opportunity costs.<sup>8</sup>

2. A second rule of thumb could be used in those cases where specialized skills are crucial for the production of a commodity and these skills are difficult to obtain in a reasonable period of time. Such commodities would be optics and precision instruments with very low tolerance levels. Again, the first type of risk is probably very low in assigning the commodity to the country with a well-developed industry and a supply of skilled labor.

3. The third rule of thumb would apply in all other cases where it may be possible to reach specialization agreements whenever the specialization is limited to very narrow boundaries. As long as each country is permitted to produce certain types of the product—for example, a certain size of ball bearing, tractors with a certain horsepower, and so

<sup>8</sup> A much stronger case could be made, of course, in a two-commodity, two-country, two-input universe where Country I uses less labor and capital than Country II in producing  $x$  and more labor and capital in producing  $y$ . In this case, by the argument of dominance, it could easily be shown that Country I has a comparative advantage in  $x$  no matter what the relative prices of labor or capital are.



forth—the probability of achieving large gains is small, but similarly the probability of losses from any of the various types of risks is minimized. These decisions can be made on the basis of purely technological coefficients, on primitive Ohlinian factor endowment considerations, or even purely arbitrarily.

The gains to be derived from this last type of specialization are based primarily on a very special type of economies of scale. The usual concept of internal economies of scale deals with construction of optimum size plants, given the technological and economic conditions, and there is a possibility of reaping these benefits. However, the economies of scale obtained from this approach are based mainly on the possibility of gaining greater economic effectiveness by reducing the number of types of products produced in a given plant, and thereby producing large runs of each type of commodity. While there is a strong probability of some gains from this source, the existence of net positive gains is not assured in all cases. It is necessary to consider possible additional transportation costs and possible complementarities between different types of products in production, research and development, or marketing. In addition, the expected gains from the reduction in the number of products produced in a given plant are based, in part, on the assumption that all the plants will continue to operate at optimum capacities. If a plant in one country is assigned a commodity for which the demand in the CEMA area is insufficient to absorb full capacity output and it is very difficult to market this commodity in the West, the benefits from serial production may be counterbalanced by increases in costs due to lower utilization of plant. Despite these caveats, it is likely that narrow intrabranch specialization among the member countries may be feasible even within all the constraints, while broader interbranch specialization may involve all the various types of risks. It may also conflict with a desire for diversified development of each economy, with a priority given to producers' goods, or with a desire to avoid abandoning existing capacities. Interbranch specialization also requires more sophisticated economic tools which have to be based on more advanced theoretical concepts and more suitable economic data than are available.

*Joint Investment Projects.* In addition to multilateral, CEMA-wide specialization agreements, there are also important discussions among two, three, or four CEMA members on the establishment of a division of labor among themselves and on the initiation of joint investment projects. These projects are a very important means of implementing a division of labor, especially in the field of raw materials. Joint investment projects represent a very interesting special type of agreement. They are the only major source of intercountry capital movements motivated by economic considerations. While capital is undervalued in

domestic price formation, its scarcity is fully recognized at the international level. Capital is the property of the state and there are ideological limitations on interest charges in intercountry loans sufficiently high to represent true scarcity, so that capital movement is very limited. However, in the case of joint projects, capital will move because the *quid pro quo* consists of deliveries of scarce and valuable raw materials. This might be regarded as a type of barter arrangement in which capital resources and raw materials are both given a value closer to their true opportunity costs.

*Summary of Limited Regret Strategy.* The limited regret strategy enables the CEMA countries to reap the benefits of comparative advantage in those cases where the benefits appear so obvious as not to require any serious economic calculations, and in other cases to obtain the benefits of internal economies of scale.

What this approach does not provide is the possibility of reaping additional gains from a broader interbranch specialization based on comparative advantage. To any economist trained in the classical theory of international trade, these gains represent the essence of the benefits to be derived from an international division of labor. This position has been challenged by a leading Hungarian economist who suggested that, with the exception of raw materials and industries with a very high raw material content, it may not be possible to reap great benefits from an attempt at specialization according to comparative advantage.<sup>9</sup> This hypothesis, which cannot be confirmed or contradicted empirically at this point, happens to fit in very well with the ideological rejection of the comparative advantage doctrine in the CEMA countries. If this hypothesis were to be confirmed, it might yield the surprising conclusion that the absence of a satisfactory theory of division of labor and the unavailability of reliable economic data do not represent serious barriers to the achievement of a highly satisfactory division of labor. If the hypothesis were to be rejected, it would then become worth while to determine the most inhibiting obstacles to the implementation of full specialization according to comparative advantage.

The three sets of obstacles that appear to be most inhibiting are independent sovereignties, insistence on diversified development, and inability to use actual or shadow prices and costs as guides for planners' decisions. It is not likely that these obstacles will be removed in the near future. Therefore, it is probable that the considerable amount of organizational and other work now going on in CEMA will not lead to the abandonment of the limited regret strategy or the achievement of a much more effective division of labor.

<sup>9</sup> Interview with Professor Istvan Friss, Director, Institute of Economics of the Hungarian Academy of Sciences, Budapest, Sept. 13, 1963

## DISCUSSION

JUDITH THORNTON: Professor Berliner has constructed several indices which he uses to compare the efficiency of the Soviet Union relative to the United States. However, it may be misleading to refer to these measures as indices of relative efficiency, since what they measure is already well known in the context of time series analysis as the difference in total productivity. Professor Berliner is using the basic technique of estimating total productivity change, and he is estimating the same variable: the unexplained residual between an index of outputs and an index of inputs. But instead of comparing a single economy at two points in time, he is comparing two economies at a single point in time.

Assuming that the two economies may be represented by a Kendrick (infinite elasticity of substitution) production function, Berliner measures the difference between the two economies as:

$$(1) \quad \frac{Q_{su}}{Q_{us}} = \frac{A_{su}}{A_{us}} \left( \alpha \frac{L_{su}}{L_{us}} + \beta \frac{K_{su}}{K_{us}} \right)$$

where  $A_{su}/A_{us}$  corresponds to Berliner's estimate of relative efficiency. Thus, if the Soviet Union is observed to produce 40 percent of U.S. output with 80 percent of U.S. input, her relative efficiency is defined as one-half of that of the U.S.

This measure of relative efficiency—actually, of relative productivity—throws little light on the question of traditional allocative efficiency. Improved resource allocation may have the same effect as improved technology on the productivity of inputs, but our measure does not allow us to separate the effects of resource allocation from other influences on the quality of inputs, including the embodiment of technology.

Turning to the indices themselves, they yield a striking conclusion: using the prices of either country, the total productivity of the measured inputs appears to be considerably lower in the Soviet Union than in the U.S. The precise amount is open to question since the results are highly sensitive to choice of price weights and since the possible margin of error in Bornstein's ruble-dollar conversion rates may be quite large.

One characteristic of the indices deserves further comment. In each of the comparisons which he makes, Professor Berliner uses opposite price weights for outputs and inputs. If the output index is weighted with American prices, then the corresponding input index is weighted with Soviet prices. The procedure is based on a recent article by Richard Moorsteen in which he argues that opposite price weights are required to assure that the inputs will be both necessary and sufficient to produce the corresponding output.<sup>1</sup>

What this procedure does, essentially, is to impose a measure of ineffi-

<sup>1</sup> Richard H. Moorsteen, "On Measuring Productive Potential and Relative Efficiency," *O.J.E.*, Aug., 1961, pp. 451-67.

ciency on each of the ratios. Given tastes and resources will generate a consistent set of equilibrium output and input price ratios. The justification for weighting with these prices is that they correspond to equilibrium marginal utilities and marginal transformation ratios. The imposition of input prices taken from another situation onto the original output prices means that we are imposing a disequilibrium onto each of the ratios. Neither set of input and output prices could be an equilibrium one for either of the situations we are comparing.

I have made a rough recomputation from Berliner's data using the same price weights for inputs and outputs.

At Dollar Values	At Ruble Values
$\frac{53.5}{.2 \times 36.4 + .8 \times 89.3} = 68\%$	$\frac{26.6}{.2 \times 26 + .8 \times 89.3} = 35\%$
	$\frac{26.6}{.4 \times 26 + .6 \times 89.3} = 42\%$

When valued at dollar values, the difference in total productivity of inputs is less marked than when valued at ruble values.

Finally, turning to the topic of static efficiency, it is worth while considering how one might measure the extent of resource misallocation in the Soviet Union. Efficient allocation implies the adjustment of all factors to underlying rates of return, net of differences in risk, so the measurement of resource misallocation involves measuring the extent to which observed factor returns exhibit persistent differences from average returns as well as estimating the amount of reallocation which would be required to bring these returns to equality.

Since capital is a more homogeneous input than labor, one approach is to measure differences in the return to capital and the extent to which the allocation of capital is responding to these differences in rates of return. The main problem in interpreting returns to capital in the Soviet Union is the existence of a disequilibrium price system combined with a system of capital rationing. An alternative approach which avoids the price problem is to measure the extent to which the allocation of capital adjusts to some other measure of excess demand for capital such as an index of double and triple shifting of machinery. These measures of allocative efficiency would provide a useful supplement to Professor Berliner's interesting data on factor productivity, and they would provide some basis for separating the effects of resource allocation from the effects of technology.

EVSEY D. DOMAR: I agree with Professor Berliner that "in the discussion of economic efficiency, all roads lead back to Pareto," though I would add that empirically they do not get us far. I hope there are not many areas in economics where the gap between refined theoretical constructs and empirical results is so wide. A study of comparative efficiency in the aggregate is a thankless task usually arousing more blame than praise. I am very glad I did not have to do it myself.

Professor Berliner is concerned with the static comparison of efficiency, while Professor Balassa reports on the dynamic one. They use, however, the same concept to measure both—a variation of Kendrick's index of total productivity—except that Berliner gives us the Soviet-U.S. ratio at a point of time, while Balassa compares the corresponding rates of growth over time. You are, of course, aware that this now popular index excludes nonphysical inputs, such as investment in health, education and training, and in research. Also, that it is simply an index of the average productivity of a combination of traditional factors of production in which the heaviest weight is usually assigned to labor, so that there is a high correlation between the movements of the Kendrick-type index and that of productivity of labor only.

Our first question is whether such an index or, more exactly, whether the rate of growth of such an index—since my assignment deals with Balassa's paper only—can serve as a reasonable approximation to what economists have meant, explicitly or not, by "dynamic efficiency." (Of course, we could settle the issue by definition, but that would be no fun.) To answer this question we have to know something about Soviet objectives and production possibilities. Such a requirement, however, is apt to put an end to any discussion of efficiency. So let us ask for a statement of Soviet objectives only. Professor Balassa provides none, but since the only result of economic activity considered by him is the rate of growth of national product, I think it is fair to interpret that as the only Soviet objective to be considered, and I am willing to go along. (A more complex welfare function would merely weaken his case.) Let us call a policy "efficient" if it speeds up the growth of output.

Now it is highly probable that many measures increasing the rate of growth of our index will accelerate the growth of output and hence be "efficient" in this sense. But this need not always be the case. How frequently we warn our students that the maximization of labor productivity does not necessarily maximize either profits or output, and this dictum holds true, though to a lesser extent, regarding the rate of growth of our index as well. Indeed, the behavior of average productivity of one or of several factors plays a modest role in optimization theory.

We could readily suggest to the Russians several measures to increase their efficiency in the Balassa sense, though I doubt their acceptance. If they reduced the rate of growth of their capital stock by shipping their old equipment to China from time to time, the rate of growth of our index would probably rise in both countries, but at the expense of Russian output. The Russian index might have grown faster if they had not expanded the cultivation on marginal lands (since no allowance for land quality is usually made in computing the index), even if they had less to eat. Both results could also be obtained by reducing female participation in their labor force, though to qualify for "dynamic efficiency" this should be done not once-and-for-all but gradually. A better method and more acceptable to this audience, but with similar results for the Russians, would be the requirement of a university degree from every entrant into the labor force.

•

Finally, it is highly probable that had the Russians been in less hurry industrializing their country, they would have received higher grades from Professor Balassa for efficiency, though they might have lost the German war. (I do not intend to carry this point too far, however. A more orderly and less hurried industrialization might have even accelerated the growth of output.)

On the whole, Professor Balassa takes a rather somber view of Soviet dynamic efficiency. True enough, based on 1937 weights, the Soviet index behaved very badly during the 1928-40 period. (With given-year weights it performed handsomely throughout, except in 1940-50; but I prefer the 1937 weights.) This and other information about that early period suggests that Soviet industrialization at the time was of a cannon-fodder variety, to borrow a military expression: huge inputs, both of labor and of capital, were thrown into the industrialization machine with rather sad results for productivity. But what else could that country have done if its population was still nearly 50 percent illiterate in 1928, if over 80 percent of its labor force was in agriculture with a minimum of mechanical skills, and if its supply of engineers and business administrators, never abundant, had been further depleted by the Civil War and emigration? For economic development, its principal asset consisted of a strong government capable of mobilizing large resources for capital accumulation, and this asset was fully used. Surely, this is exactly what our own theory of comparative advantage as applied to industrialization would call efficient. This is not to say that the Russians did everything right. Their record abounds in appalling stupidities (such as the forced collectivization of the peasants), but the poor behavior of the index, taken by itself, is no proof of inefficiency in the sense that their output would have grown faster if some other method of industrialization had been used. For that matter, Kendrick's own index shows a much slower growth in the United States before World War I than after. The sharply reduced immigration must have been at least partly responsible for the subsequent acceleration of the growth of the index, but was it "inefficient" of the U.S. to permit a virtually unrestricted immigration prior to World War I?

In the more recent period of 1950-58 the Soviet index certainly blossomed up by any standards, growing at roughly the same annual rate as the German and the Japanese, and much above those of the U.S., Canada and the U.K., to mention only the countries referred to in the paper. Perhaps the recent performance of the index is more significant than its rate of growth over the whole 1928-58 period which was broken up by the German invasion. Professor Balassa might have at least recognized this fact in following Bergson's practice in counting only twenty-six rather than thirty years between 1928 and 1958 for computational purposes, though this is hardly a sufficient adjustment since Soviet prewar output was not reached until about 1948.

All this sounds rather critical of Professor Balassa's paper, though my only real quarrel with him is not about his paper but about its purpose and title. It is indeed an interesting paper, and had he entitled it "A Compara-

tive Study of Sources of Economic Growth in the Soviet Union and the United States" I would have had few objections. He might have then proceeded along Edward Denison's lines even if he made as many assumptions as Denison did or more.

With this peaceful (to Professor Balassa) suggestion, I intended to end my comments, if I had not been provoked by the penultimate sentence in his paper. It reads: "However, in terms of dynamic efficiency, the performance of the Soviet economy does not appear to be superior to that of free enterprise economies at a similar level of economic development." Professor Balassa may be perfectly right, but how many free enterprise economies appeared in his paper? (No other evidence was cited.) The Russians must be indeed flattered that in judging their performance we point to the best among capitalist countries. What about the run-of-the-mill, such as Poland or Rumania in the interwar period or Latin America today? And besides, how can one come to such a momentous conclusion on the basis of the behavior of one crude index?

FREDERIC L. PRYOR: Dr. Neuberger has given us some interesting insights into the inefficient economic integration<sup>1</sup> of the European Communist nations.

The "minimum regret approach," which bears the brunt of his analysis of the course of economic integration in the East bloc, is essentially a decision-making model which focuses attention on certain benefits and risks of integration, viewed from the perspective of specified political and economic policy objectives and restraints. Although he does not attempt to place some quantitative values to these risks and benefits, he has implicitly recognized that such an approach runs two related dangers. First, it is difficult to use such a model to predict which economic sectors will be integrated in the future. Second, the integration which is occurring in any sector can be justified by simply asserting that the net benefits of integration were great enough to impel the Council for Economic Mutual Assistance (CEMA) to act.

In order to avoid these dangers Dr. Neuberger has derived several rules of thumb for decision making from his models which suggest certain sectors where integration should proceed the fastest. He has placed particular emphasis on the relative ease of integrating the raw materials sector—an hypothesis which can be statistically tested.

Any kind of sophisticated hypothesis testing is difficult because few bloc nations have published sufficiently detailed trade statistics for any length of time. One simple test can be made by examining the changes in the trade-production ratios for various commodities. The greater the increase in these trade-production ratios, the greater the degree of integration activi-

<sup>1</sup> He has implicitly assumed that the integration is inefficient and, although aggregative measurements are difficult to make, all evidence points in this direction. For instance, one recent study, based on a regression analysis of the foreign trade of Western nations, showed that the trade volume of each Communist nation is less than 50 percent that of a hypothetical capitalist nation with a similar population and per capita production. Calculations from my *The Communist Foreign Trade System* (Cambridge, Mass., and London, England, 1963), p. 27.

ties. Since not enough information is available to construct such ratios for many commodities for the bloc as a whole, it is necessary to look at such ratios for a single country. Moreover, in order to provide some basis of comparison, it is necessary to look at the changes of these ratios in time periods before and after the integration work began in earnest. Since more such trade-production ratios over an extended time period can be calculated for East Germany than for any other CEMA nation (excluding the U.S.S.R.), it is perhaps best to use East Germany to test Neuberger's hypothesis.

The average change and the standard deviation of change of trade-production ratios of commodities in the raw materials, producer goods and consumer goods sectors (about 170 commodities in all) in East Germany have been estimated for the time periods 1955-58 and 1958-61.<sup>2</sup> According to these calculations the changes in the average trade-production ratios and in the standard deviation of change of these ratios were greatest in the producer goods sector, followed by the consumer goods sector. Because of the crudeness of the data and the statistical methods employed, Neuberger's hypothesis that integration should occur most rapidly in the raw materials sector is not conclusively disproven; nevertheless, some doubt is cast on his model.

Of course, Neuberger also suggested that there could be a rapid specialization of production for those producer goods for which each nation could produce different models or sizes of the goods. Nevertheless, by interpreting his rules of thumb of CEMA decision making more loosely, we run the danger of being able to justify any statistical results.

The major problem in setting up any model of decision making in the CEMA is that we know too little about the actual procedures of this organization. If we are ever to learn anything concrete about the way in which the CEMA and the individual countries evaluate the benefits and risks of economic integration, it must come through a thorough analysis of the available trade statistics; i.e., through the method of "revealed preferences."

If Neuberger's hypothesis is incorrect, it is useful to try to uncover the underlying reasons. The most important factor in the slowness of economic integration in the raw materials sector lies, I believe, in the system of foreign trade prices used in intrabloc trade.

Since 1954, all East bloc nations have used approximate "capitalist world market prices" in their trade with each other. If any other prices had been used, this might have encouraged certain bloc countries, which felt they were paying or receiving disadvantageous prices, to trade with the West. This use of such capitalist prices in intrabloc trade has discouraged an all-bloc division of labor in the following way:

Bloc country X, which is producing a good for which it feels it has a comparative disadvantage vis-à-vis, the world, is tempted to stop producing the good and import it from either the West or another bloc country. How-

<sup>2</sup> *Ibid.*, p. 39.



ever, if this country has a comparative advantage in producing this good vis-à-vis the other bloc countries and if the CEMA policy-makers feel that this good should not be imported from the West, then it would be to the bloc's advantage as a whole that country X produce this good. Nevertheless, country X might not want to specialize in the production of this good unless it could receive a special price over and above the capitalist world market price or unless it would be able to import certain goods which it is currently producing and in which it has an even greater comparative disadvantage vis-à-vis the world. With the present bloc methods of estimating the profitability of foreign trade of various commodities, it might be quite difficult to conclusively demonstrate the savings possibilities of the latter alternative. This problem has been recognized in some of the more recent East bloc foreign trade literature. The situation is further complicated by the fact that while the national planners of some bloc nations think in terms of comparative advantage, most of the decision making within the CEMA appears to be in terms of absolute rather than comparative advantage.

The dilemma can be stated in a more general way: Any set of sovereign nations intent on economic integration must develop its own set of enforceable intrabloc prices, different from the prices of the outside world, if this economic integration is to proceed very far. Extensive smuggling would defeat such an arrangement. Because the central authorities of the CEMA do not have the power to prevent any nation disadvantaged under such a new intrabloc price system from trading with the outside world (smuggling by the government) it is very difficult for the CEMA to develop its own price system. In other words, the CEMA does not have the functional equivalent of an enforced all-bloc tariff wall. The limits of space prevent a deeper examination of this restraint on decision making in the CEMA.

Dr. Neuberger's judgment about the absolute belief of all bloc policy-makers on the benefits of the development of all major sectors of the national economy and the necessity for the priority development of the producers' goods industries in each country is open to question. The economic literature of the various bloc countries suggests a slowly growing sophistication within some East bloc nations on matters of comparative advantage and foreign trade theory.

INVITED DISSERTATIONS, I

THE MONETARY MECHANISM: SOME PARTIAL  
RELATIONSHIPS\*

By JAMES L. PIERCE  
*Yale University*

The existence of an intimate connection between monetary policy and the commercial banking system is well known. The exact nature of this connection is far from perfectly understood, however. The purpose of this paper is to outline an attempt to specify both the nature and the degree of control that the monetary authorities can exert over commercial banks. Such a specification involves the formulation of the response mechanism of banks to policy. While such a specification is admittedly a small part of the total reaction of the economy to monetary policy, it is an important one; and it represents a logical first step in the determination of this monetary mechanism. The primary goal of this study is the determination of the role of policy; the portfolio analysis which is used is a means to this end. The discussion which follows will be carried out in terms of an individual commercial bank.

The bank is conceived to have a total supply of funds at its disposal comprised of its deposit liabilities (less required reserves) and of its capital account. Given these funds, the bank distributes them among available assets. The total asset portfolio is aggregated into three groups: a transactions balance, a portfolio of relatively long-term bonds, and a group of nonfinancial loans. The transactions balance will be called the portfolio of reserve assets; it provides the bank with a pool of highly liquid assets to be used for transactions purposes. The components of the portfolio are the bank's holdings of cash, Treasury bills, and other highly liquid assets. The portfolio of relatively long-term securities will be called the investment portfolio; it is held for income and diversification purposes and it provides a source of speculation. This portfolio contains such securities as intermediate and long-term government bonds, municipal bonds, and special long-term issues. The nonfinancial loan portfolio is held for income purposes and is composed of all loans other than the extremely short-term loans made to brokers, dealers, and finance companies; these are included in the reserve asset group. The following notation may be used:

\* I should like to thank James Friedman for reading earlier versions of this paper and for making many extremely valuable suggestions. Donald Hester and James Tobin were kind enough to bring some important omissions to my attention. I also wish to thank the Board of Governors of the Federal Reserve System for providing access to needed materials. Part of the research for this paper was supported by the National Science Foundation.

A: Total assets (less required reserves)  
 R: Reserve assets  
 I: Investment assets  
 L: Loan assets  
 F: Total available funds  
 D: Demand deposits (less required reserves)  
 T: Time deposits (less required reserves)  
 C: Capital accounts

These terms may be arranged in the following balance-sheet identities:

$$A = R + I + L$$

$$F = D + T + C$$

$$F = A.$$

The analysis of portfolio selection is most easily handled by assuming that each asset group is composed of a homogeneous set of assets; i.e., all the assets within any group are perfect substitutes. The reserve, investment, and loan groups will be respectively characterized by cash, long-term bonds, and loans. Each asset has its own return and risk properties. The portfolio problem involves an analysis of how a bank will allocate a given dollar of  $F$  among the three assets in a manner which will maximize its expected utility.

Given the demand nature of many of its liabilities and the small size of the capital account relative to total liabilities, it is safe to assume that the bank is a risk averting investor. Under the further assumption that the bank considers its estimate of the standard deviation of asset return to be the relevant measure of risk, its preferences are approximated by a quadratic utility function.<sup>1</sup>

The bank is not able to completely and reliably schedule its transactions needs; this being the case,  $R$  must also serve as a buffer stock which insulates the other asset groups from unexpected variations in transactions. A clearer isolation of the factors which condition the bank's demand for reserve assets may be achieved by making two initial simplifying assumptions. It is assumed that demand deposits are the bank's only deposit liability and that the level of loan demand is known and constant.

As a means of handling the expected volume of transactions, it is assumed that this volume is a positive function of the bank's deposit size; and it is further assumed for simplicity that the relationship is proportional. The desired reserve asset portfolio will be expressed in terms of the expected value and standard deviation of the future level of deposits where the standard deviation is used by the bank as a measure of the risk that actual deposits will differ from their forecasted or expected value. Assume that the bank has  $N$  depositors, each with a given

<sup>1</sup> Harry Markowitz, *Portfolio Selection* (John Wiley & Sons, 1959), Chap. 10.

size deposit at the beginning of any period. Further assume that the total level of deposits at the end of the period is a random, lognormally distributed variable with expected value  $\alpha$  and variance  $\beta^2$ . The lognormal distribution has several characteristics which make it particularly well suited for the analysis at hand.<sup>2</sup> Consider a random variable,  $x$ , defined over the range  $0 < x < \infty$ . If  $y = \log(x)$  is normally distributed with mean  $\mu$  and variance  $\sigma^2$ , then  $x$  is lognormally distributed. The mean of  $x$ ,  $\alpha$ , is given by  $\alpha = e^{\mu + 1/2\sigma^2}$ , and the variance  $\beta^2$ , by  $\beta^2 = e^{2\mu + \sigma^2}(e^{\sigma^2} - 1) = \alpha^2\eta^2$  where  $\eta^2 = (e^{\sigma^2} - 1)$ , and  $\eta$  is the coefficient of variation of the distribution. The distribution of  $x$  is completely specified by the parameters  $\mu$  and  $\sigma^2$ .

Abstracting from seasonal and other systematic influences, it is assumed that the bank uses the current level of deposits as a measure of the deposit level which it expects to prevail at the end of the period; i.e.,  $E(x_{t+1}) = \alpha_{t+1} = x_t$  and therefore,  $E(y_{t+1}) = \mu_t$  where  $x$  denotes the total level of deposits. It is further assumed that the bank considers the parameter  $\sigma^2$  to be a constant for all  $t$ . This assumption implies that given percentage deviations of  $x$  from its expected value are viewed by the bank as having the same likelihood of occurrence no matter what the current level of  $x$  (or of  $\mu$ ). It can be seen from the expression above that  $\eta$  depends only on  $\sigma^2$  and, hence, is a constant under this assumption. It is also seen that the standard deviation of  $x$  is given by  $\beta = \alpha\eta$ , so  $\beta_{t+1} = \alpha_{t+1}\eta = x_t\eta$ . The bank thus considers the standard deviation of  $x$  in period  $t+1$  to be proportional to the current level of deposits,  $x_t$ .

The proportionality of  $\beta_{t+1}$  to  $x_t$  rests on the assumption that the number of depositors,  $N$ , is fixed. In general, however, an increase in current deposit size will at least in part represent an increase in the number of depositors. If these new accounts are not perfectly correlated with the old, or with each other, the standard deviation of deposits will not rise in proportion to the increase in total deposit size. The extent to which the bank considers  $\beta$  to change will depend upon the degree to which the movements in deposits represent a change in the number of depositors and the extent to which it represents an alteration in the size of existing accounts. Only if variations in total deposits arise solely from this latter source will total deposit risk vary proportionally with deposit size; such a phenomenon is indeed unlikely to occur.

Analyses of the transactions demand for money have indicated that the transactions balance should vary positively but less than proportionally with the volume of transactions.<sup>3</sup> For similar reasons, the de-

<sup>2</sup> J. Aitchison and J. A. C. Brown, *The Lognormal Distribution* (Cambridge Univ. Press, 1957), pp. 7-19.

<sup>3</sup> Cf. James Tobin, "The Interest Elasticity of Transactions Demand for Cash," *Rev. of Econ. and Statis.*, Aug., 1956, pp. 241-47, and William J. Baumol, "The Transactions Demand for Cash: An Inventory Theoretic Approach," *Q.J.E.*, Nov., 1952, pp. 545-56.

sired reserve asset portfolio will also vary positively but less than proportionally with expected deposit size. The portion of  $R$  which represents a safety allowance will likewise vary less than proportionally with the level of deposits. Given that the total supply of available funds,  $F$ , is composed almost exclusively of deposits, the desired reserve asset portfolio will be a positive, less than proportional function of this total supply.

The composition of  $F$  should have an important influence on the size of desired  $R$ . The larger the proportion of highly active and/or erratic deposits in the total supply of funds, the larger should be the desired reserve asset portfolio, given  $F$ . Such deposits tend to increase both the expected volume of transactions and the variability of this volume.

The entire discussion of the reserve asset portfolio has implicitly assumed the existence of alternative uses of funds. The deposit forecast and the confidence which the bank places in it do not uniquely determine the size of  $R$ . Given the size and composition of  $F$ , desired  $R$  will vary negatively with the rates of return on investment and loan assets. The desired reserve asset portfolio will vary positively with the transactions costs involved in the purchase and sale of competing assets, and with the risks associated with holding the other two portfolio groups; i.e., risk of default and/or of capital loss. Finally, the  $R$  which the bank wishes to hold will vary positively with the penalties involved in being caught with insufficient reserve assets to meet a deposit loss; such penalties include the costs of borrowing and the expenses involved in making frequent short-term adjustments in the investment asset portfolio.

The discussion has so far been conducted under the assumptions that the level of loan demand and the yield on investment assets are known and constant. The characteristics of  $I$  and  $L$  must now be briefly discussed.

The investment asset portfolio is intermediate with respect to the other two portfolio groups in terms of expected return, liquidity, and risk. The expected return on  $I$  is higher than for  $R$ , but such return can be gained only at the cost of decreased portfolio liquidity and increased asset risk. While investment assets are marketable, they do not provide the high liquidity of short-term assets. The existence of relatively high brokerage fees, of random variations in price, and of possible market "thinness" make them ill-suited for the short-run manipulations required of reserve assets. Rapid and frequent movements into and out of the portfolio in response to unexpected deposit movements would tend to reduce the net yield on these assets below the yield on  $R$ . Brokerage fees and short-term price fluctuations made the net yield on  $I$  in part a function of the length of time these assets are held. The

holding period for investment assets is too long to make them suitable for transactions purposes. If the bank has to sell investment assets to meet a deposit loss, such sale is possible, however; this gives  $I$  an important advantage over loan assets which, of course, are not marketable.

The existence of fairly large variations in the prices of investment assets makes such securities subject to risk of capital loss. In deciding how to best allocate a given dollar of funds between  $R$  and  $I$ , the bank must compare expected transactions needs against the expected return and risk on  $I$ . The actual allocation which is made will depend upon the shape of the bank's utility function. Under the assumption that the function is quadratic in return, the higher the expected return or the lower the risk on  $I$ , the greater will be the share of investment assets in a dollar of available funds, given expected transactions needs.

The relative shares of the reserve and investment asset portfolios in the total asset portfolio are conditioned by the bank's decision to hold loan assets. Of the three asset groups, loans have the highest expected rate of return, the greatest risk, and the least liquidity. Of the three asset groups, only loans possess significant risk of default. Due to the relatively small capital account of a typical commercial bank, such risk is of crucial importance. The expected return on loans is, of course, conditioned by the interest rate charged by the bank. Such return tends to exceed the expected return on the other two asset groups.

A relatively high rate of return and comparatively large risk provide loans with their two essential characteristics. Under the assumption of the quadratic utility function, the share of total available funds allocated to loans will vary positively with expected return and negatively with loan risk. Unlike the other two asset groups, the return and risk on loans are subject to some control by the bank. The bank will be faced by a demand for its loans which is a decreasing function of the interest charge on loans. Further, if it is assumed that the bank considers such factors as loan size, maturity, and guarantees as important elements in the determination of loan risk, these factors are also subject to bank manipulation. Other things being equal, borrowers prefer relatively large average loan size, long maturity, and no guarantees; banks on the other hand prefer relatively small loan size (per loan), short maturity, and strong guarantees. The work of Hester and of Guttentag indicates that it is useful to combine these return and risk factors into what Hester calls a set of loan terms.<sup>4</sup> If stringent loan terms are associated with relatively high interest rates, short maturities, small average loan size, and strong guarantees, the amount of available funds which

<sup>4</sup> Donald D. Hester, "An Empirical Examination of a Commercial Bank Loan Offer Function," *Yale Economic Essays*, Vol. 2, No. 1, pp. 3-57, and especially pp. 3-12, and Jack Guttentag, "Credit Availability, Interest Rates, and Monetary Policy," *So. Econ. J.*, Jan., 1960, pp. 219-28.

the bank wants to put into loans will be positively related to the stringency of loan terms and the demand for loans will be negatively related to this stringency.

It is likely that the supply of loans function is nonlinear. Given the bank's aversion to risk and the total supply of funds at its disposal, the bank will require ever increasing stringency of loan terms to induce it to supply a constant increment to the loan portfolio. Given  $F$ , the bank can expand  $L$  only at the expense of  $R$  and  $I$ . Such expansion increases the risk associated with unexpected deposit losses and it increases portfolio default risk. The bank's relatively small capital account coupled to ever present transactions needs should be sufficient to produce a less than proportional response of the quantity of loans supplied to a given change in loan terms.

If information on loan terms were generally available, the discussion would now be complete. It could be argued that the supply of loans function is accurately specified by  $F$ , its composition, the yield on  $I$ , and the stringency of loan terms. Loan demand, on the other hand, depends upon many factors other than the stringency of loan terms—factors which are often much more important to the borrower (e.g., expected future profits) than the terms on which he can obtain loans from the bank. The demand function is far from being specified. An observed relationship between loan terms and  $L$  would be the locus of equilibria between the demand and supply functions. This locus would approximate the bank's supply function if the variance of the stochastic term of the supply function were small compared to that of the demand function. This condition would be more than adequately met in the case at hand. The unexplained variance of the supply function would certainly be much smaller than the unexplained variance of the demand function, which has been taken to be only a function of the stringency of loan terms. This being the case, a more complete specification of the demand for loans function would be unnecessary for identification of the supply schedule.<sup>5</sup> Unfortunately, data on loan terms are not available. A proxy variable must be used to represent the relationship between loan terms and  $L$ .

In a world in which loan demand is subject to cyclical variation, the bank needs some measure to tell it how its current portfolio position compares to "normal." Normal in this case refers to the cycle average of the proportion of  $F$  devoted to loans. The bank compares the present share of  $L$  in  $F$  to its average share; i.e., the bank uses the variable  $(L/F - \bar{L}/F)$ , where  $\bar{L}/F$  is the cycle average of  $L/F$ , as an indicator of the desirability of its current portfolio position. It is assumed that in a cyclical setting, the bank expects a movement in loan demand in

<sup>5</sup> E. J. Working, "What Do Statistical 'Demand Curves' Show?" *Readings in Price Theory* (Richard D. Irwin, 1952), pp. 97-118.

one direction to be followed by another movement in the same direction. With this expectation and given the current value of  $(L/F - \hat{L}/F)$ , the bank must decide by how much it is willing to alter  $L$  for a given expected change in loan demand. It is argued that the larger  $(L/F - \hat{L}/F)$ , the less the bank will respond to a given change in loan demand. Thus,  $L$  should vary positively but less than proportionally with  $(L/F - \hat{L}/F)$ . It should be noted that if the deviations of  $L/F$  from its cycle average were replaced by a measure of the stringency of loan terms, one would expect  $L$  to bear precisely the same sort of relationship to loan terms.

The argument must be altered when discussing the case of cyclical contractions in loan demand. Bond prices, as well as loan demand, experience cyclical variation. The two cyclical patterns should be roughly coincident. At the upper turning point in loan demand, bond prices tend to be low (high interest rates). As loan demand begins to fall, the bank has an incentive, not to greatly reduce loan terms, but rather to let  $L$  decline and to put the liberated funds into a bond market which is expected to rise. As the decline in loan demand continues and as bond prices rise, the incentive to undertake further bond price speculation will be reduced as the bank expects both loan demand and bond prices to reverse their direction of movement in the near future. The bank will begin to realize its gains from the investment portfolio and to build up its reserve asset position for the expected rise in loan demand. Within the context of the short-term cyclical contractions in GNP which have characterized the postwar period, the realization of bank expectations during periods of declining loan demand should be fairly complete.

Holding all other supply determining factors constant,  $L_t$  will depend upon the value of  $(L/F - \hat{L}/F)_{t-1}$  and upon the direction in which loan demand is moving. The relationship may be expressed in the following manner: define a variable  $X_e$  which has a value of unity for contractions in loan demand and a value of zero for expansions. Leaving out other factors for the moment, the desired relationship may be expressed by  $L_t = a_1 + (a_2 + a_3 X_e)(L/F - \hat{L}/F)_{t-1}$ . The coefficient of  $(L/F - \hat{L}/F)_{t-1}$  will thus depend upon whether loan demand is rising or falling.

The total model may now be presented. The proportion of available fund devoted to loans is estimated using the following relationship:<sup>6</sup>

$$\begin{aligned} \log \frac{L_t}{F_{t-1}} = & a_1 + [a_2 + a_3(X_e)_t] \log \left[ \left( \frac{L}{F} - \frac{\hat{L}}{F} \right)_{t-1} + 5.0\% \right] \\ & + a_4 \log \left( \frac{D_0}{F} \right)_{t-1} + a_5 \log F_{t-1} + a_6 \log i_{t-1} + u_t \end{aligned}$$

<sup>6</sup> All ratio forms are expressed as percentages. The addition of the arbitrary constant to  $(L/F - \hat{L}/F)_{t-1}$  renders this variable strictly positive. The coefficients  $a_2$  and  $a_3$  are not unique with respect to choice of the additive constant.



where  $D_g$  is government deposits (which are particularly unstable),  $i$  is the yield on 3-5 year government bonds, and  $u$  is the stochastic term. The structural relationship between  $L$  and  $F$  is assumed to take the form:  $L = F^\gamma$ , thus  $L/F = F^{(\gamma-1)}$ ;  $L/F$  is conditioned by the size of  $F$ . The coefficient  $a_3$  is an estimate of the structural parameter  $\gamma$  less unity. The discussion of the relationship between  $R$  and  $F$  implies that  $\gamma > 1$ .

The data used to test the model were obtained from the balance sheets of eighty-five large commercial banks, from all over the country, which are particularly active in the money market. The data are in the form of time series aggregates composed of weekly observation of the banks' portfolios covering the period from July, 1959, through December, 1962. The results are as follows:

$$\log \frac{L_t}{F_{t-1}} = 1.577 + \left[ \begin{array}{ccc} .071 & + & .001(X_e)_t \\ 5.9 & 22.2 & 5.9 \end{array} \right] \log \left[ \left( \frac{L}{F} - \frac{\hat{L}}{F} \right)_{t-1} + 5.0\% \right]$$

$$- .010 \log \left( \frac{D_g}{F} \right)_{t-1} - .108 \log (F_{t-1}) - .063 \log (i)_{t-1}$$

$$- 5.7 \qquad \qquad - 5.3 \qquad \qquad - 6.09$$

$$R^2 = .914 \qquad D = 1.7$$

The corresponding  $t$ -statistics appear below each parameter estimate.

Given the nature of the data and the simplicity of the model, the fit is surprisingly good. The intensity and the direction of movement of loan demand have an important influence on the loan portfolio. The coefficient for  $(L/F - \hat{L}/F)_{t-1}$  strongly implies that  $L/F$  responds less than proportionally to the stringency of loan terms. The elasticity of  $L/F$  with respect to this proxy for loan terms is significantly less than unity. The coefficient is also significantly larger during contractions in loan demand than during expansions. The bank is more willing to reduce the proportion of  $L$  in  $F$  for a 1 percent decline in  $(L/F - \hat{L}/F)$  than it is to increase it for a 1 percent rise. Such reaction implies that the banks have incentives to channel funds liberated by declining loan demand into the bond market.

The variable  $D_g/F$  has been used because government deposits in commercial banks tend to be highly unstable.  $D_g$  was the only unstable deposit item which could be isolated from the data. The influence of government deposits on the loan account is relatively important. A 1 percent rise in  $D_g/F$  will induce the banks to reduce the proportion of  $L$  in total assets by .01 percent. The composition of  $F$  is an important element in the portfolio decision.

The coefficient of  $F$  needs some explaining. It has been argued that  $a_3$  would exceed unity; the discussion of the relationship between  $R$  and

$F$  has implied that  $\gamma - 1 > 0$ . The negative elasticity coefficient for  $F$  implies that a 1 percent increase in  $F$  will produce a .89 percent increase in  $L$ . The original contention that  $\gamma$  exceeds unity was based on several assumptions regarding the relationship between  $R$  and  $F$ . It has been assumed that banks are willing and able to completely adjust their desired loan portfolios to variations in  $F$  with a one period lag. If, in fact, the lag is longer than this—perhaps distributed through time—the coefficient  $a_8$  will tend to lie below its “equilibrium” value. The possibility of such a lag deserves further study. Only if it proves possible to estimate the lag will it be possible to evaluate the validity of the assumptions made in deriving the relationship between  $L_t$  and  $F_{t-1}$ , which lay behind the assertion that  $\gamma > 1$ . Until this is done, little can be said concerning the economies or diseconomies of scale exhibited by the banks in the sample.

Finally the elasticity coefficient for  $i$  is quite interesting. This coefficient indicates that the investment portfolio is an important competitor for bank funds. Irrespective of the direction of movement of loan demand, a 1 per cent rise in the yield on 3–5 year government securities produces a .06 percent decline in  $L/F$ . Loan demand movements do not swamp all other portfolio considerations.

In the introduction to this paper it was stated that the purpose of the study was to specify the mechanism of response of commercial banks to monetary policy. On the basis of the sample used in the empirical work, some conclusions regarding this component of the monetary mechanism are made. In the interest of brevity, the only policy variable to be considered is  $F$ . Through its open market operations and its control over reserve requirements, the Federal Reserve System can determine the supply of available funds within tolerably narrow limits for the banking system as a whole. The manner in which the banks allocate these funds among alternative asset groups depends upon the strength and direction of movement of loan demand, the existing composition of  $F$ , and upon the yield on investment assets. The influence which variations in the total supply of funds exert on the banks' loan portfolios is a matter of predictable economic choice. The empirical evidence strongly suggests that the size of  $F$  has had, and will have, an extremely important influence on the size of  $L$ . The extent of the influence will depend upon the factors discussed. Banks do not blindly meter out a constant proportion of  $F$  to the three portfolio groups; and they do not passively react to variations in loan demand. The link between policy and loans is a strong one; but  $F$  is only one factor among many. If the quantity of loans is a relevant target variable for policy, the target can be hit only if these other factors are explicitly and carefully taken into account in the policy decision.

# THE TERM STRUCTURE OF INTEREST RATES\*

By BURTON G. MALKIEL  
*Princeton University*

## *Introduction and Background*

Recent discussions of the term structure of interest rates have been trenchant and discordant. Many participants have been attracted toward unicausal explanations of the rate structure. Meiselman [2] holds that the traditional expectations theory can provide a complete description of the behavior of default-free yields over the current century. Kessel [3] also accords a crucial role to expectations, but finds the Hicksian liquidity-premium variant to be the appropriate formulation of the model. Culbertson [5] [6], on the other hand, has come close to contending that the short- and long-term debt markets are essentially independent and finds the expectational explanation theoretically unsatisfactory and devoid of empirical relevance. The apparent dichotomy of views was most sharply defined at last year's meetings of the Econometric Society. Here Kessel's advocacy of the Hicksian liquidity-premium model met the disfavor of both Meiselman and Culbertson. Moreover, Culbertson chided his colleagues for seeking "casual correlations" rather than "causes." He claimed "the Lutz expectational theory as applied to long-term interest rates scarcely merits a second glance, as it is obviously inconsistent with all we know about the way people do behave in debt markets, and the way they are able to behave. . . ."

It is my view that, apart from aesthetics, monistic explanations of the determination of the rate structure are unsatisfactory both theoretically and empirically. In this paper, I shall first summarize briefly some of the points made in my dissertation that bear directly on the controversy noted above. Then I shall present in as much detail as space will permit some of my empirical and theoretical work which attempts to reconcile and synthesize opposing hypotheses.

In the first place it should be noted that, contrary to Culbertson's view, it is possible to formulate a version of the expectations theory that both accords with the practices of bond investors and offers added insights into the behavior of the yield curve [see 1]. Moreover, my variant of the traditional theory lends itself well to empirical testing. The results of such tests support the modified expectations hypothesis.<sup>1</sup>

\* A full list of the people who have helped me in this study would deplete the space allotted to this paper. Nevertheless, I must at the very least express my profound gratitude to my dissertation committee: W. J. Baumol, L. V. Chandler and R. E. Quandt; and to E. J. Kane for countless hours of discussion, criticism, and counsel. I am also thankful to the National Science Foundation whose support helped in the completion of this paper.

<sup>1</sup> In performing an empirical test, I examined whether the behavior of the interest rate struc-

I also find that the introduction of transactions costs affords further insights into the term structure. I offer here, without the evidence, my major conclusions. To the extent that bond issuers require funds over a period longer than the holding periods of bond investors, the yield curve will be positively sloped (when rates are expected to remain unchanged), with the effects of transactions costs felt mainly in the very early maturities. I maintain this is the typical case since most firms require relatively permanent financing. The empirical data on transactions costs also provide a convincing case that government bond dealers are Hicksian risk averters. Finally, I find that the transactions costs of issuing debt securities provide a rational explanation for the maturity sectoring of bond issuers over a wide range of interest rate differentials.

*Institutional versus Expectational Theories: Some Evidence on  
the Microeconomic Level*

I turn now to the major purpose of this paper: the empirical investigation of the microeconomic evidence that bears upon the competing hypotheses and the construction of a theoretical model that synthesizes these positions. By the "institutional theory" of the rate structure I shall mean the extreme form of the argument which holds that risk aversion and institutional impediments so restrict the mobility of market participants that the long- and short-term markets are essentially independent. Interest rates in each market are then determined solely by their internal supply and demand interactions since neither borrowers nor lenders allegedly shift between markets in response to rate differentials. On the other hand, the pure expectations theory implies that, unless they alter expectations of future rates, relative changes in the supply of bonds cannot influence the term structure. Meiselman [2, p. 57] has suggested that market-excess-demand schedules for securities of each maturity tend to be infinitely elastic, despite the presence of individual participants who speculate or hedge on the basis of risk aversion. He holds that this results from the overlapping of transactors who specialize in particular maturity ranges and from the presence of adequately-financed speculators who are willing to switch from market to

---

ture was consistent with my basic expectational model. The model determines the slope of the yield curve solely on the basis of expectations of a "normal range" of interest rates in conjunction with the mathematical relationships inherent in bond price movements. I assumed that investors were agnostic regarding the direction of future rate changes but that they firmly believed that any interest rate variations would be contained within the normal range. Investors were assumed to shape their expectations of the normal range by adding a specified number of standard deviations to either side of the average of rates over a period of past years. The slope of the yield curve was then postulated to be a function of the relationship between the current level of rates and the bounds of the normal range as suggested by my model. The results (utilizing the Durand data) supported the modified expectations hypothesis. They provide independent evidence corroborating the findings of Meiselman [2] and Kessel [3] which affirm the importance of expectations.

market whenever one offers terms even slightly more favorable than the other. They thereby adjust the quantities of securities taken from or supplied to each market in a manner which maintains a rate structure consistent with their expectations of future short rates.

It is convenient to distinguish between two kinds of speculation. A corporation portfolio manager, with funds to invest for six months who buys a five-year bond, anticipating a larger holding period yield thereon, is engaged in a speculative transaction. He gives up a certain six-months return for an uncertain one. Such activity (portfolio flexibility) is essential to the expectations theory. The extreme institutionalists deny its existence. A second class of speculative activity we shall call "professional speculation." Its essential characteristic is the simultaneous maintenance of long and short positions in different maturities. If a dealer sells short a six-month certificate and simultaneously buys a five-year bond, this is considered professional speculation. Temporarily, I shall ignore this latter activity.

I first investigated whether data on portfolio practices of institutional investors lent credence to the extreme institutional hypothesis of portfolio behavior. Post-accord statistics<sup>2</sup> on institutional holdings of governments revealed no evidence of complete segmentation. On the contrary, the maturity distribution of the aggregate portfolios of these institutions has been remarkably changeable. Institutions with predominantly short-term liabilities have substituted broadly from short through intermediate issues (up to five years and slightly longer), while institutions with long-term liabilities have shifted from long through intermediate issues (down to five years and slightly shorter).<sup>3</sup>

There are, however, several difficulties with such an aggregate approach. In the first place, only movement can be measured, whereas a measure of mobility is desired. While mobility can always be inferred from movement, absence of movement does not imply lack of mobility. Actual movement will be frustrated to the extent the rate structure already reflects investors' evaluations of the appropriate yield relationships. Moreover, the changing composition of reporting institutions and the lack of detailed breakdowns in the early periods mar the usefulness of the *Survey* for long-run comparisons. Finally, since the *Survey* covers only governments and since the composition of the liabilities of the reporting institutions is unknown, any inferences must be treated cautiously. Consequently, I felt it was imperative to investigate the portfolio behavior of these institutions on the microeconomic level before drawing conclusions about the competing hypotheses.

I interviewed officers of several money market banks, of a sample of

<sup>2</sup> From the Treasury Survey of Ownership [4].

<sup>3</sup> During 1959 and 1960 life insurance companies (classically long-term investors) actually purchased significant quantities of governments with maturities under five years.

other financial institutions who are important participants in the debt markets, and a number of government bond dealers. I reviewed each institution's portfolio behavior during the past decade with reference to the composition of its assets and liabilities and its responses to changes in expectations. This examination indicated that my inferences from the aggregate, cross-sectional data did not overstate the degree of flexibility characterizing the practices of portfolio managers. In fact, I found repeated examples of expectations-induced shifts in portfolio composition of a much greater magnitude than appears in the aggregate data.<sup>4</sup>

I conclude from this investigation that it is unreasonable to talk of segmentation in any absolute sense.<sup>5</sup> However, the fact that some institutionalists may overstate their case should not obscure the important kernel of truth in their position. Various market participants do have maturity preferences. This was stressed repeatedly during my interviews. We must reinterpret segmentation to mean simply that many buyers and sellers must be paid differential premiums to induce them to move from their preferred maturities. The bribes necessary to induce investors who prefer short maturities to move to the longest maturities may indeed be larger than the constellation of interest rate spreads encountered in practice. Consequently such financial institutions may never have been induced to buy significant quantities of long-term bonds. Nevertheless, we reject that no differentials exist sufficient to induce any movement. The evidence suggests that sufficient differentials have frequently existed to induce many investors to shift readily over broad maturity ranges. These investors provide the links in a continuous chain tying together all maturity sectors of the bond market.

I conducted a similar investigation of the supply side of the market. The institutionalists are correct in asserting that, within the relevant range of rate differentials, private and municipal bond issuers have not been induced to make long-run changes in the maturity composition of their debt. But what they overlook is the degree of flexibility available to issuers to adjust the timing of long-term bond issues. While transactions costs may prevent issuers who need funds over a long period from borrowing perpetually at short term, these costs will not prevent some borrowing at short term for a limited period until conditions improve

<sup>4</sup> For example, one commercial bank during 1958 had a portfolio in which over two-thirds of the securities were to mature within one year. A year later, the composition of the portfolio had been changed so that almost two-thirds of the securities had maturities longer than one year with a significant portion maturing in from three to five years. The shift was explained entirely by a change in expectations.

<sup>5</sup> Even where I discovered that risk aversion or other considerations led to legal or conventional sanctions prohibiting the purchase of issues in certain maturity ranges, they did not prevent substitutability over broad maturity ranges of the yield curve.

for long-term borrowing. Using aggregate data and the results of extensive interviews, I find considerable evidence that expectations strongly influence the timing of many long-term bond issues.

While my investigation indicated that an unqualified version of the institutional theory of the rate structure is unacceptable, it suggested that the pure expectations theory was equally extreme. In particular, I found Meiselman's reconciliation inapplicable in two respects. Either demanders or suppliers must be free of risk aversion and have uniform expectations or there must exist a group of adequately financed professional speculators who render the excess-demand schedules for securities of different maturities infinitely elastic at rates consistent with their (uniform) expectations. My empirical investigation convinces me that neither of these assumptions reflects the actual workings of the market.

In the first place, I have already suggested that maturity preferences affect the behavior of participants on both sides of the market. Moreover, my attempt at a direct examination of the extent to which expectations were uniformly held led me to take a more Keynesian view. Keynes [8, Chap. XV] characterized uncertainty as the existence of a dispersion of opinions constituting the full range of "bull-bear" positions. One of my tests for uniformity of opinions was to examine the market letters of four leading bond investment advisors [10] [11] [12] [13] over the period 1958-61. I assumed that these widely read letters, whether they simply reflect or actually form existing expectations, can serve as a surrogate for questioning the actual market participants. While my examination was largely impressionistic, it proved impossible to resist the conclusion that the bond market is typically characterized by an enormous dispersion of opinion. Even when the letter writers (infrequently) concurred on the direction of rate changes, there was still wide disagreement on the magnitude of rate movements.

Empirical investigation of the actual workings of the market also convinces me that serious impediments delimit the volume of professional speculative transactions. First, there is the borrowing cost of  $\frac{1}{2}$  of 1 percent incurred on securities sold short. Moreover, institutional restraints limit the maturity of the issues that government bond dealers and other speculators may use to borrow short-term funds on repurchase agreements. In the case of dealer banks, complications arise from the income tax laws which encourage banks to make each year either a loss or gain year. Perhaps most important, it is enormously difficult to sell short significant quantities of governments because of the complexities (and, in many cases, the impossibility) of borrowing issues.<sup>6</sup> Finally, whatever speculative transactions are possible are not free of transac-

<sup>6</sup> In this connection, I recommend that the Treasury make securities from the Trust Funds available for borrowing. This would vastly improve the functioning of the market.

tions costs, and it is very likely that professional speculators have diverse expectations and demand Hicksian liquidity premiums. Since government bond dealers (who, because of their information and expertise, are most likely to speculate professionally) charge additional fees to take positions in longer-term issues in the course of their regular business, it seems reasonable to expect that they would demand the same sort of risk premium protection when they assume positions as speculators. Therefore, such limited professional speculative activity as does occur is not likely to be conducted in terms of expectations alone, and it is not structured to lead to long rates which are any simple geometric average of expected short rates.

*A Model to Synthesize the Expectations and Institutional Theories*

I then construct a model to determine the effect of diverse expectations, limited supply flexibility, and maturity preferences. While space limitations preclude a description of the full model, diagrammatic illustrations of some of the two-market cases will offer the flavor of the analysis. In the tradition of rate structure theory, my treatment is one of partial equilibrium and comparative statistics where the level of interest rates and the total demand and supply of funds are taken as data. The question I ask is what rate structure will make investors and issuers satisfied with their relative holdings (issues) of the different maturities of debt. Supply and demand schedules are constructed which take account of existing expectations and whatever inflexibilities and maturity preferences are assumed. The experiment is then repeated with a different maturity distribution of inflexible suppliers,<sup>7</sup> and the equilibria are compared. I treat three distinct cases representing different combinations of maturity preferences and diversity of expectations.

Case 1: *Diverse Expectations—No Maturity Preferences.* Figure 1 treats a case with only two securities, a one-year and a two-year issue.<sup>8</sup> The box diagram indicates a fixed total volume of funds. Investors are assumed to have no preferences between maturities but to have diverse expectations. No professional speculators are assumed to exist.  $CC'$  is the curve giving the percentage of holdings demanded in the form of two-year (and hence residually, of one-year) bonds as a function of the interest rate differential. It is constructed via the standard Hicks-Lutz averaging formula from the density function below that arrays investable funds according to their owners' firm expectations of the forward one-year rate. For example, point  $C$  on the demand function corresponds to point  $c$  on the density function. This is the threshold interest rate

<sup>7</sup> By inflexible suppliers we mean those whose supply schedules of bonds of different maturities are completely inelastic to interest rate differentials over the relevant range.

<sup>8</sup> The notation follows that utilized by Meiselman [2]. Capital  $R$ 's represent actual rates; small  $r$ 's, expected rates. The prescript gives the period when the rate becomes applicable; the subscript, the duration of the loan.



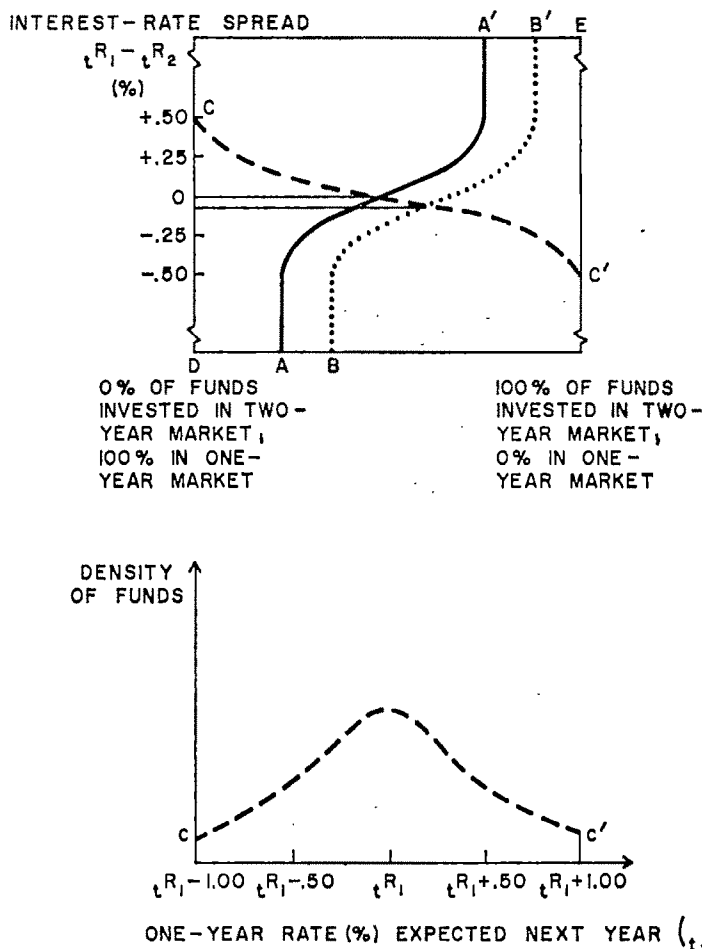


FIGURE 1

differential at which the first dollar of funds is attracted into the two-year market. Investors who expect that the forward (one-year) rate will be 100 basis points (1.00 percent) below the current (one-year) rate (point  $c$  above the axis in the lower figure) will enter the two-year market whenever the two-year rate is no more than (approximately) 50 basis points below the one-year rate (point  $C$  in the upper diagram).<sup>9</sup> At point  $C'$ , corresponding to  $c'$ , the last dollar of funds enters the two-year market. Here, relative interest rates induce all investors to invest

<sup>9</sup> The diagram assumes the current one-year rate to be 3.5 percent. By the averaging formula, unity plus the two-year rate,  $1 + {}_tR_2$ , is the geometric mean of (unity plus) the current one-year rate (1.035) and (unity plus) the anticipated forward one-year rate, which, at point  $c$ , is 2.5 percent. Thus we have  $1 + {}_tR_2 = \sqrt{(1.025)(1.035)} = 1.03$  (approx.). Hence the interest rate spread,  ${}_tR_1 - {}_tR_2$ , is .035 - .030 = .0050 which gives us the level of  $C$  in the upper graph.

in long maturities. At intermediate interest rate spreads, just those investors who feel that the actual rate on two-year bonds is higher (lower) than the rate called for by their expectations purchase two-year (one-year) bonds.

$AA'$  is the supply curve of bonds. We assume  $DA$  of two-year bonds and  $A'E$  of one-year issues are supplied irrespective of the interest rate spread.<sup>10</sup> All remaining bond suppliers will vary their maturity with relative interest rates. The interest-elastic segment of the supply curve assumes diverse expectations on the part of suppliers and its construction is completely analogous to that of the demand curve. As drawn,  $AA'$  and  $CC'$  intersect at a zero interest rate spread that happens to be consistent with the mean expectations of suppliers and demanders. But note that any change in the amount of bonds issued by inflexible suppliers will upset this relationship. For example, if inflexible suppliers choose to move  $AB = A'B'$  bonds from the short to the long market, the relevant supply curve becomes  $BB'$  and the one-year rate must then lie below the two-year rate. It can also be shown that, over the bulk of the relevant range, the greater the variance of  $cc'$  (i.e., the greater the diversity of expectations), the greater will be the slope of our demand curve, and therefore the more sensitive will the rate structure be to such supply changes.<sup>11</sup>

**Case 2: Uniform Expectations—Maturity Preferences.** We next treat a case where all market participants hold uniform expectations that the future one-year rate will equal today's one-year rate.<sup>12</sup> In this case, however, we assume that different market participants have definite preferences and must be paid to be induced to shift from their preferred maturities. In Figure 2, the discontinuous demand function  $FGHF'$  assumes that funds  $FG$  belong to investors with strong preferences for income certainty (over the two-year period) who must be paid a premium of 50 basis points on one-year issues to induce them to invest in these securities. Funds  $HF'$  belong to investors with a strong preference for principal certainty (i.e., they are Hicksian risk averters)

<sup>10</sup> While this analysis does not include transactions costs explicitly, we have suggested earlier that issue costs may lead to a sectoring of bond suppliers over a wide range of interest rate differentials. The complete inelasticity of this portion of the supply is taken to hold only over the relevant ranges of interest rate differentials.

<sup>11</sup> On the other hand, if all market participants hold identical expectations and have no maturity preferences, then  $CC'$  becomes infinitely elastic and relative supplies of securities become irrelevant. This is the limiting case of the expectations hypothesis which Meiselman treats. We should also note that if a significant group of investors hold similar expectations, some small changes of supplies may have little effect despite the wide range of existing expectations. Needless to say, the very limited amount of professional speculation which I find can easily be accommodated by the model. But to the extent that these speculators have diverse expectations and some degree of risk aversion as I have suggested, our results will not be changed.

<sup>12</sup> Little generality is lost by assuming no change in anticipated forward rates. The effect of an expected rise or fall is readily taken into account by means of the averaging mechanism.

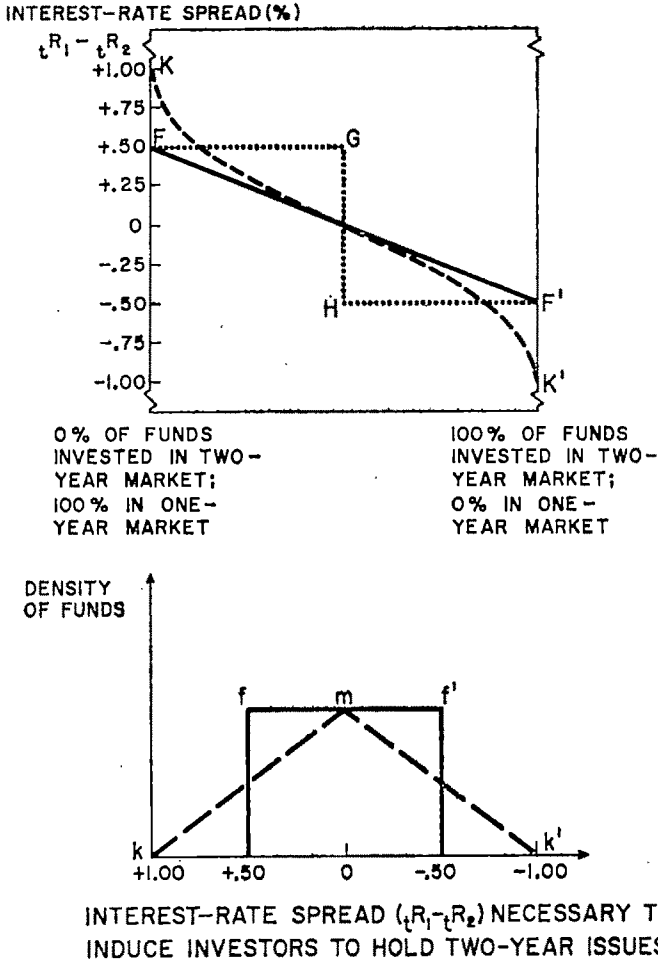


FIGURE 2

who demand a 50 basis-point premium on two-year issues to induce them to enter the two-year market.

The alternative demand curve  $FF'$  may be a more realistic construction in light of the indication of my interviews that the strength of such maturity preferences differs among and within classes of financial institutions.  $FF'$  is derived from the rectangular density function  $ff'$  which now arrays investable funds according to the interest rate differential necessary to induce them into the two-year market. The area under the density function to the left of any point on the horizontal axis represents the fraction of investors who want to invest in two-year issues at the rate differential indicated at that point. Because the density function of Figure 2 already has these interest rate differentials on the

horizontal axis, the construction of the demand curve is now trivial. For any rate spread along the horizontal axis of the density function and the area to the left of that point, the corresponding point on the demand function (upper figure) has as its ordinate the same rate differential and, as its abscissa, the above-mentioned area.

In general, the slope and curvature of  $FF'$  depends upon the distribution of funds among different financial institutions and of the degree of risk aversion among portfolio managers. While no supply function is drawn on Figure 2, we could derive one analogous to  $AA'$  in Figure 1. The interest-elastic portion would, in this case, represent issuers' different maturity preferences. Again we would find that the term structure is sensitive to relative supply changes despite uniformity of expectations.

*Case 3: Diverse Expectations—Maturity Preferences.* The most interesting, and empirically the most relevant, case assumes that market participants have diverse expectations and maturity preferences. We will first treat the simple variant (similar to Case 1) where all participants have diverse expectations but the same maturity preference. For example, let us add to Case 1 the assumption that all investors are to the same degree Hicksian risk averters. Here, we would add a constant Hicksian liquidity premium ( $L_h$ ) to the anticipated forward rate. The density function  $cc'$  would shift to the right and the demand function  $CC'$  would shift downward. This increases the probability that the yield curve will be positively sloped; i.e., that backwardation will be present. Similarly, if all demanders are taken to prefer certainty of income over the long run and, hence, longer issues, the probability of a contango is increased.

The complicated case treats both diverse preferences and expectations. Let us assume (as in Case 2) that the density function  $ff'$  characterizes the maturity preferences of the market. We assume further that an identical rectangular distribution describes the subjectively appropriate yield differentials justified by the diverse expectations of market participants.<sup>13</sup> Positing the two distributions to be independent, we may form the combined density function to account for the composite effects of subjectively appropriate differentials arising out of maturity preferences and diverse expectations. The combined density function for simple distributions may be found by the method of convolution.<sup>14</sup> In Figure 2, density function  $kmk'$  is found by convolving two rectangular distributions of the form  $ff'$ . The distribution  $kmk'$  arrays each

<sup>13</sup> In terms of the density function of Figure 1, we assume the presence of a rectangular distribution of expectations of next year's forward rate ( ${}_t+1r_1$ ) over the range  ${}_tR_1 - 1.00 \leq {}_t+1r_1 \leq {}_tR_1 + 1.00$ .

<sup>14</sup> The density function of the sum of two independent continuous random variables is the convolution of the density functions of the random variables. See [15, Chap. 20].

dollar of funds in the market by the total interest rate spread necessary to induce investment in the two-year issue, taking account of both expectations and maturity preferences. From the combined density function, we may then construct the demand function  $KK'$  directly. We note the range of  $KK'$  is wider and its slope is nowhere less steep than that of  $FF'$ . Thus, the presence of both diverse expectations and maturity preferences will usually increase the influence of supply changes on the term structure. But, a model which accounts for these influences does not vitiate the importance of expectations. On the contrary, expectations remain a major force in the determination of the rate structure and, as expectations at any time become more uniform, the term structure becomes less sensitive to changes in the relative supply of debt instruments.

In a three-market variant of this model I have easily accounted for some humps occasionally observed in empirical yield curves. In a situation where the yield curve would otherwise be flat, a nonexpectations-induced increase in the supply of intermediate-term issues produces a humped curve.<sup>15</sup> I submit that this construction best explains the humped yield curves in the government securities market during 1959 and 1960. The increase from mid-1958 to mid-1960 in the supply of bonds with maturities between three and five years (the location of the hump) was one of the most significant postwar changes in the maturity composition of the federal debt.

### *Conclusion*

In conclusion, my study reaffirms the importance of expectations in the determination of the term structure, although I favor a formulation of the theory which differs somewhat from the Lutzian [9] model. Moreover, I disagree with both extreme views on the matter—the one which assigns to expectations the unique and complete role in determining the rate structure and the other, which offers them no role at all. My view is essentially syncretic and lies closest to that suggested by Conard [7]. I believe that emendations must be made to the expectations analysis to account for transactions costs, diversity of expectations, and institutional maturity preferences on both sides of the market. These suggest that supply levels do affect the term structure. Much more empirical work is needed to establish the degree of sensitivity of the rate structure to relative supply changes. My theoretical and empirical findings on the microeconomic<sup>16</sup> level, however, lead me to

<sup>15</sup> A pure expectations theory can offer only a rather clumsy explanation of humps in intermediate maturities of the yield curve.

<sup>16</sup> The question of the sensitivity of the term structure is by no means settled by an analysis of aggregative data alone. A careful recent study by Wood [14], however, finds that changes in private and government securities outstanding do help to explain variations in the short-long spread.

believe that this sensitivity may be greater than many expectationists believe.

## REFERENCES

1. Burton G. Malkiel, "Expectations, Bond Prices and the Term Structure of Interest Rates," *Q.J.E.*, May, 1962, pp. 197-218.
2. David Meiselman, *The Term Structure of Interest Rates* (Prentice-Hall, 1962).
3. Reuben Kessel, *The Cyclical Behavior of the Term Structure of Interest Rates* (N.B.E.R., Dec., 1962).
4. United States Treasury, *Treasury Bull.* (1952-63).
5. J. M. Culbertson, "The Term Structure of Interest Rates," *Q.J.E.*, Nov., 1957, pp. 485-517.
6. ———, "Discussion: Econometric Studies in Money Markets II" (a paper presented to the Econometric Society, Pittsburgh, Dec., 1962).
7. Joseph W. Conard, *Introduction to the Theory of Interest* (Berkeley: Univ. of California Press, 1959), Part Three.
8. John Maynard Keynes, *A Treatise on Money*, Vol. 1 (London: Macmillan, 1930).
9. Friedrich A. Lutz, "The Structure of Interest Rates," *Q.J.E.*, Nov., 1940, pp. 36-63.
10. *Moody's Bond Survey* (1958-61).
11. Sylvia Porter, *Reporting on Governments* (1958-61).
12. T. Goldsmith, *Goldsmith's Washington Service* (1958-61).
13. Murray Olyphant, "Government Bonds," *Banking*, 1958-61.
14. John H. Wood, "An Econometric Model of the Term Structure of Interest Rates" (a paper presented to the Econometric Society, Pittsburgh, Dec., 1962).
15. Harold Freeman, *Introduction to Statistical Inference* (Addison-Wesley, 1963).

# COMMERCIAL BANK PORTFOLIO ADJUSTMENTS\*

By WILLIAM R. RUSSELL  
*University of Wisconsin*

## I. Introduction

The model of commercial bank portfolio adjustments presented here consists essentially of two parts. First, a static model is formulated. This is developed along the same lines as earlier works of J. Tobin,<sup>1</sup> H. Markowitz,<sup>2</sup> and D. Farrar.<sup>3</sup> The static model shows the "desired" or "equilibrium" portfolio the banker would choose under given economic conditions. Second, the model is expanded to include a mechanism which shows how the portfolio adjustments take place over time. This part of the model is a slight modification of an earlier model presented by L. Telser.<sup>4</sup>

## II. Maximizing in an Environment of Uncertainty<sup>5</sup>

In microeconomic theory it is commonly assumed that each decision-making unit maximizes its utility. Whereas certainty is assumed in microeconomic theory, it is uncertainty which presents the major problem for the portfolio manager and may not be assumed away. Hence rather than maximizing actual utility, the banker (decision-maker or collective "mind" of the decision-making committee) is assumed to maximize expected utility.<sup>6</sup>

*A View of the Market—Prior Subjective Probability Distribution of Prices of Assets.* In making investment decisions, the rational banker must ultimately "guess" what annual payments and prices will be in the future for securities under consideration for investment. But since expectations with respect to annual payments tend to be reflected in the expectations of prices, no great distortion would be expected by

\* The original version of this paper was a Ph.D. thesis submitted to the University of Washington. I am especially grateful for criticism and suggestions received from Arnold Zellner and Walter Oi which materially improved this study. My thanks also to Yoram Barzel, James Crutchfield, Donald Gordon, David Huang, Jan Kmenta, Thomas Saving and Ernst Stromsdorfer for their helpful comments. I claim all remaining errors. Finally I would like to gratefully acknowledge that a large part of this study was made possible by the generous support of the Social Systems Research Institute.

<sup>1</sup> James Tobin, "Liquidity Preference As Behavior Towards Risk," *Rev. of Econ. Studies*, Feb., 1958.

<sup>2</sup> Harry Markowitz, *Portfolio Selection*, Cowles Foundation, Monograph No. 16, 1959.

<sup>3</sup> Donald Farrar, *The Investment Decision Under Uncertainty* (Prentice-Hall, 1962).

<sup>4</sup> Lester Telser, "The Demand for Branded Goods as Estimated From Consumer Panel Data," *Rev. of Econ. and Statis.*, Aug., 1962.

<sup>5</sup> The development of this section draws heavily from Harry Markowitz, *op. cit.*, and Donald Farrar, *op. cit.*

<sup>6</sup> For a justification of selecting the maximization of expected utility see Markowitz, *op. cit.*, pp. 205-42 or Duncan R. Luce and Howard Raiffa, *Games and Decision* (John Wiley & Sons, 1958).

suppressing the specific variations in payments expectations.<sup>7</sup> Consequently the annual payments of a security will be taken to be certain for the decision period. Suppose all information held by the banker about sets of asset characteristics pertinent to the banker's evaluation of securities is translated into expected future price behavior in the form of subjective probability distributions (SPD) of prices. Each set of characteristics describes an "abstract" asset. The future prices of these abstract assets are considered random variables whose SPD are determined by the banker's evaluation of the probability of occurrence of each price at some future time  $t$ . The SPD of prices for each abstract asset is assumed to be constant over time. But while the SPD of prices of these assets are fixed, those for "market" assets are not. A bond maturing, for example, continually changes from one abstract class to another until at maturity it "enters" that abstract class with the set of characteristics describing cash. Or if in the opinion of the banker the quality of a market asset deteriorates, it becomes a "different" asset, entering a different abstract asset class which would likely have a SPD with a higher variance.<sup>8</sup>

*Evaluation of Portfolios—the Optimal Portfolio.*<sup>9</sup> The portfolio is defined as a linear combination of the holdings of individual assets. Since the prices of assets are given by SPD the value of any portfolio at some future point in time will also be given by a SPD. Since each abstract asset has in general a different SPD, the SPD of the value of the portfolio changes as different compositions are considered. It is now necessary to determine how the various compositions will be evaluated for selection. This evaluation depends on the utility function of the banker. It will be denoted by:

- (2.1)  $U = U(Z_t)$  the certainty utility function with utility a function of  $Z_t$ , the terminal value of the portfolio.

The terminal value of the portfolio at time  $t$  is defined as the value of the portfolio at time  $t$  plus the unplanned algebraic net change in other accounts of the banker's balance sheet. The changes in most of the liabilities and capital accounts are known in advance and to a great extent controlled by the banker. These known changes can be considered as affecting the total value of the portfolio at time 0. However among the bank's liabilities are the demand and time deposits which are due on demand.<sup>10</sup> The banker does not know with certainty the net

<sup>7</sup> Take the case of a perpetual corporation bond. If the probability of a delinquent interest payment rises an investor would re-evaluate his price expectations for that asset and in general would conclude the mean of the future prices of that bond would be lower.

<sup>8</sup> The variance may be used as an indication of risk. See Harry Markowitz, *op. cit.*, p. 286.

<sup>9</sup> This development parallels Donald Farrar, *op. cit.*

<sup>10</sup> Time deposits are not technically due on demand but banks typically honor these liabilities on demand



changes of these liabilities for they are essentially determined by the actions of depositors. This introduces a variable largely independent of the banker's control. Suppose this change in total deposits at time  $t$  is taken as a random variable and totally independent of the banker's control. Introducing this variable explicitly into the utility function the result is:

$$(2.2) \quad U = U(A_{T,t} + N_t)$$

where

$A_{T,t}$  = the value of the current total assets at time  $t$ . It is considered a random variable with mean  $\bar{A}_{T,t}$  and variance  $s_A^2$ .

$N_t$  = the net change in total deposits at time  $t$ , considered a random variable with mean,  $\bar{N}_t$  and variance  $s_N^2$ .

If the utility function is assumed to be quadratic,<sup>11</sup> the expected utility can be shown to be given by:

$$(2.3) \quad E(U) = U(\bar{A}_{T,t} + \bar{N}_t) + \frac{s_A^2 + s_N^2}{2} U''(\bar{A}_{T,t} + \bar{N}_t).$$

The expected utility is a function of the expected value of the portfolio, the expected changes in total deposits, the variance of the value of the portfolio, and the variance of changes in total deposits.

The maximization of expected utility with respect to the amount of each asset to be held in the portfolio, constrained by the total funds available to be invested at time 0, yields the optimal portfolio. This optimal set of assets,  $\hat{A}_1, \hat{A}_2, \dots, \hat{A}_n$ , will be a function of expected changes in deposits and current prices of assets in the market. By a transformation of the variables the optimal set of assets may instead be given as a function of the current market rates of interest. That is.

$$(2.4) \quad \{A_1, A_2, \dots, A_n\} = \{h_1(i_1, i_2, \dots, i_n, \bar{N}_t), \\ h_2(i_1, i_2, \dots, i_n, \bar{N}_t) \dots \\ h_n(i_1, i_2, \dots, i_n, \bar{N}_t)\}.$$

The optimal portfolio depends therefore on the structure of rates prevailing in the market and the expected change in deposits. This model will be further expanded to describe portfolio adjustments toward equilibrium.

### III. Portfolio Adjustments Over Time<sup>12</sup>

At equilibrium the asset composition held in the commercial bank portfolio will be the desired or optimal composition. A change, for

<sup>11</sup> For a discussion of further implications of this assumption see K. Arrow, "Comment: The Portfolio Approach," *Rev. of Econ. and Statis.*, Part 2, Supplement, Feb., 1963.

<sup>12</sup> The extension of the model in this section relies on a model developed by L. Telser, *ibid.*

example, in current interest rates changes the optimal portfolio, making it necessary for the commercial banks to rearrange their portfolios. But this adjustment may take one period or a number of periods. That is, a given structure of rates may have to be maintained for many periods before the equilibrium portfolio is attained.

Suppose one asset is considered. If all other assets are considered collectively as one alternative asset class to asset  $i$ , the expected amounts of asset  $i$  may be given by:

$$(3.1) \quad A_{i,t} = A_{i,t-1}P_{ii} + A_{a,t-1}P_{ai}$$

where

$A_{i,t}$  = amount of asset  $i$  at time  $t$

$A_{a,t-1}$  = amount of "other" assets held at time  $t-1$

$P_{ii}$  = transition probability from  $A_i$  to  $A_i$

$P_{ai}$  = transition probability from  $A_a$  to  $A_i$

The probability that a banker will retain or switch funds is not in general fixed. The earlier section suggests that the probability varies as interest rates or expected changes in deposits vary. If these transition probabilities are given as functions of the market rates of interest and expected changes in deposits, the expected proportion of any asset  $i$  in a portfolio may be given by:

$$(3.2) \quad \begin{aligned} A_{i,t} = & A_{i,t-1}P_{ii}(i_1, i_2, \dots, i_n, \bar{N}_t) \\ & + (1 - A_{i,t-1})P_{ai}(i_1, i_2, \dots, i_n, \bar{N}_t). \end{aligned}$$

The basic hypothesis this study poses is that banks seek to maintain a certain asset structure in a given economic environment. This desired asset structure is functionally related to the market interest rates and to expected changes in deposits (positive or negative) at time  $t$ . If the desired portfolio deviates from the actual portfolio, bankers adjust the distribution of assets from period to period to make the current portfolio approach the desired. The model suggests that as the market rate of interest of an asset rises, *ceteris paribus*, the portfolio share held in that asset also rises. If a negative change of deposits, i.e., a drain of funds is expected, *ceteris paribus*, the percent of the portfolio held in cash will increase and the shares held in earning assets will decrease.<sup>12</sup>

#### IV. Specification of the Model

The model could be estimated with each return individually entering the transition probability functions as an independent variable. Since

<sup>12</sup> Actually an increase in short-term assets or other relatively low variance assets as an outflow is expected is not strictly inconsistent with the model. But if the expectations refer to a short decision period (assumed to be a month in this study), it is more likely that the percent of short-term assets decreases and the proportion of cash in the portfolio increases.

these market rates of interest tend to move together a variable was constructed which was in the form of the difference between the return of the asset considered and that for the "other" asset. The greater the positive difference, the higher  $P_{ii}$  and  $P_{oi}$ . The return for "other" assets is a weighted average<sup>14</sup> of the market interest rates for the alternative assets considered. Taking the linear form for the transition probability functions, the following are obtained:

$$(4.1) \quad P_{oi} = a_{oi} + b_{oi}i_{ii,t}^* + c_{oi}\bar{N}_t$$

$$(4.2) \quad P_{ii} = a_{ii} + b_{ii}i_{ii,t}^* + c_{ii}\bar{N}_t$$

where

$i_{ii,t}^*$  = the difference between the market rates of returns of asset  $i$  and "other" assets.

Assuming equal structural parameters for all banks, and independence of the distributions of lagged percent asset shares and expected percent changes in deposits, the actual aggregate<sup>15</sup> portfolio percentage of asset  $i$  is given by:

$$(4.3) \quad \begin{aligned} S_{i,t} = & a_{oi} + b_{oi}i_{ii,t}^* + c_{oi}D_t + (a_{ii} - a_{oi})S_{i,t-1} \\ & + (b_{ii} - b_{oi})(S_{i,t-1})i_{ii,t}^* + (c_{ii} - c_{oi})(S_{i,t-1})D_t + v \end{aligned}$$

where

$S_{i,t}$  = aggregate commercial banks portfolio share of asset  $i$  at time  $t$

$D_t$  = aggregate percent change in total deposits

$v$  = random error term.

Since it is likely that a variable which is the product of two variables is closely correlated with its factors, an alternative form of equation (4.3) is investigated. Taking linear approximations<sup>16</sup> for each of the products,  $S_{i,t-1}(i_{ii,t}^*)$ , and  $S_{i,t-1}(D_t)$ , and substituting, equation (4.3) reduces to:

$$(4.4) \quad S_{i,t} = B_0 + B_1S_{i,t-1} + B_2i_{ii,t}^* + B_3D_t + v.$$

The coefficients are related to transition probabilities functions parameters as follows:

<sup>14</sup> The weights were assumed to follow a geometrically declining one parameter distribution. The value of the parameter yielding the "best fit," i.e., highest coefficient of determination, was chosen. The weights were applied to the returns on the basis of a priori ranking of the assets according to variance.

<sup>15</sup> Although I would have preferred to test this model at the individual bank level, it was impossible to obtain such data. Hence it was necessary to aggregate.

<sup>16</sup> For discussion on the use of linear approximation for products see Lawrence Klein, *Econometrics* (Row, Peterson & Co., 1956), pp. 120-21 and Lester Telser, *op. cit.*, pp. 310-11.

$$B_0 = a_{ai} - (b_{ii} - b_{ai})(av S_{i,t-1})(av i_{i,t}^*) - (c_{ii} - c_{ai})(av S_{i,t-1}) av D_t$$

$$B_1 = (a_{ii} - a_{ai}) + (b_{ii} - b_{ai}) av i_{i,t}^* + (c_{ii} - c_{ai}) av D_t$$

$$B_2 = b_{ai} + (b_{ii} - b_{ai}) av S_{i,t-1}$$

$$B_3 = c_{ai} + (c_{ii} - c_{ai}) av S_{i,t-1}$$

where

$$av S_{i,t-1} = \text{average } S_{i,t-1}, \quad av i_{i,t}^* = \text{average } i_{i,t}^*, \quad av D_t = \text{average } D_t.$$

Unfortunately the model is now underidentified. The parameters of the transition probabilities functions cannot be obtained from estimates of the  $B$ 's. But by rearrangement and appropriate substitution, the following relations are obtained:

$$av P_{ii} - av P_{ai} = B_1$$

$$av P_{ai} = B_0 + B_2 av i_{i,t}^* + B_3 av D_t.$$

Since

$$(4.5) \quad P_{aa} + P_{ai} = 1$$

$$(4.6) \quad P_{ii} + P_{ia} = 1$$

all the transition probabilities can be evaluated at the mean values of the variables.

*The Data.* The data used in this study consisted of monthly observations over the period from June, 1947, to December, 1961, reported in the *Federal Reserve Bulletins*. The portfolio of the commercial banks was limited to those of cash, U.S. government securities, and bank loans. The asset cash consisted of cash in vault, deposit reserves with the Federal Reserve banks, and deposits in other banks.<sup>17</sup> The government securities recorded at par values were divided into three classes: the short, intermediate, and long terms. The short-term government securities were those securities maturing or callable within one year. The intermediate terms were government securities maturing or callable between one and five years of date. The long terms were government securities maturing or callable after ten years of current date. The asset loans were defined as being equivalent to the Federal Reserve definition of total loans. This classification included loans with significant differences in the maturity dates and terms of loans. Moreover, it included loans from every source, combining consumer loans, business loans,

<sup>17</sup> It might be pointed out at this point that no account has been taken of the reserve requirements. Actually the reserve requirement places a minimum constraint on the asset defined as cash. Therefore bankers may be in disequilibrium and unable to adjust toward the desired portfolio.

farm loans, etc. Since adequate breakdowns were not available for monthly periods, this total class was used.

The rate of return of cash was taken to be zero. The three-month Treasury bill rate was used as the rate of interest for the short terms. The average rate on government securities maturing between three-five years was used as the rate of interest for the intermediate term asset class. The average rate of government securities maturing after ten years was used as the market rate of return for the long terms. The prime loans rate was used as the market rate of return on loans.

The expected changes in deposits were taken to be a function of the actual percentage change in total deposits over the next future period; i.e., equal to  $(\text{total deposits}_{t+1} - \text{total deposits}_t)$  divided by the total deposits<sub>t</sub>. A negative value would correspond to an outflow of funds. To the extent that commercial banks can forecast accurately the changes in total deposits within the period, the actual changes would tend to equal the expected.<sup>18</sup>

None of the variables were seasonally adjusted. The banker looking to his portfolio position surely would anticipate seasonal factors and operate accordingly. To remove the seasonal elements would remove some of the reason for the observed portfolio.

*The Empirical Results.* Equation (4.4) was estimated for each of the five asset classes: cash, short terms, intermediate terms, long terms, and loans, by least squares. Tests for autocorrelation of the residuals in each asset equation yielded inconclusive results. The regressions results are consistent with the hypotheses proposed by this model. The coefficients of determination obtained ranged from .7475 for the short terms to .9736 for loans. The coefficients for the lagged asset market share were positive and significant at the 1 percent level for all assets. In each case this coefficient was less than 1, assuring stable equilibrium portfolio shares for all assets. The estimated coefficients for the variable representing rate of return had the predicted sign for all assets. In every case the portfolio share of an asset increases as its rate of return increases relative to the rate of return of other assets. This coefficient was significant at the 1 percent level for cash and at the 10 percent level for the long-term assets. But for short- and intermediate-term assets, the standard errors were larger than their coefficients.

The magnitude of the coefficients for the rate-of-return variable shows the responsiveness of an asset share to changes in the differential between its own rate and the rate of return of "other" assets. By this measure the assets loans and cash were the two most responsive. In general, the monthly period asset share adjustment in response to

<sup>18</sup> Neglecting distributional effects. This of course presumes fulfillment of assumptions on aggregation.

changes in the interest rate variable was small. The largest was the asset loans response of .578 percent of total portfolio. However, if no other changes occur, adjustments continue period to period until the equilibrium asset shares are held.

The absolute portfolio share responses of each asset as the market interest rate of a substitute increases 1 percent were investigated. If the interest rate on loans increases 1 percent, .28 percent of the total portfolio value would be switched from the asset cash to loans. This substitution would account for about half of the expected increase in loans. As the average rate of interest of government securities maturing over ten years increases, the major portfolio share substitution is again a decrease of the portfolio share of cash. As the average market rate of the three-five year government securities increases, the major decrease in asset shares would be that of the short terms. The major response of substitute asset market shares as a result of an increase in the three-month Treasury bill rate is the lowering of the portfolio share of loans.

Judging by the values of the cross-elasticities, the asset loans is the closest substitute to cash, intermediate terms, and bonds. These asset shares appear especially responsive to changes in the rate of return on loans.

The long-run loan elasticities ranged from a low of .609 for short terms to a high of 1.08 for long terms. Only in the cases of the long terms and loans did the elasticity exceed 1. In every instance the long-run elasticities were greater than the short-run. The bankers are able to make a more complete adjustment in the long run. Needless to say the low values obtained from the elasticities do not mean that the equilibrium shares do not change significantly. Take the case of an increase of 1 percent, from 3 percent to 4 percent, in the interest rate of loans. The share of loans would increase about .15 from its mean share of .46 of the total portfolio.

It was hypothesized that a portfolio with a higher expected return and higher variance would be desired as bankers expected unplanned addition of funds to the portfolio. The signs of the coefficients for the variable representing expected increases in total deposits were consistent with the hypothesis. The signs were positive for each of the earning assets and negative for the asset cash. However, two of the five coefficients were less than their standard errors. Only the coefficient for cash was significant at the 1 percent level.

The greatest asset share response to expected changes in total deposits was that of the short-term assets. The portfolio share response for the short terms was about sixfold the magnitudes of those for the intermediate or long terms. The loans class was least responsive to expected changes in the level of deposits. Since the expected change in total

deposits was to be over the next one-month period, it was not surprising that the major change in earning assets portfolio shares was in the short terms.

The low values obtained for the coefficients of both variables  $\pi^*_{i,t}$  and  $D_{i,t}$ , and the high coefficient for lagged shares indicate the tendency of the portfolio composition to change little in a period. The high repeat transition probabilities and low transfer probabilities obtained also illustrate this. Since a period is one month, this result was not unexpected. The aggregate percent share of any asset is not likely to change greatly in a single month. However, if the interest rates and the expected changes in deposits were to be fairly constant for many periods, a more complete adjustment could be made. The current portfolio would thus approximate more closely the desired portfolio.

*A Test of the Results.* An independent test of the coefficients was made possible by the second World War. In March, 1942, the Federal Reserve and the U.S. Treasury Department agreed to peg the interest rates at the structure then prevailing in the market. The rate structure remained pegged until early 1945. The equilibrium shares predicted by the model were calculated for this period and compared with the actual asset shares. Although not conclusive, the results indicated that the actual changes were consistent with the estimates. In the particular case of cash shares, not only did the estimated equilibrium share coincide with the apparent actual equilibrium, but the estimated time required for adjustment was surprisingly close to the actual.

### *Summary and Conclusion*

A general framework for the analysis of commercial banks' portfolio adjustments has been presented. A dynamic microeconomic model has been constructed based on the assumption of maximizing behavior. The model indicated that for any asset in the portfolio, the proportion of that asset would increase as its expected rate of return increases relative to that of other assets. The regression analysis yielded results consistent with this hypothesis. The predicted directions of asset shares response to changes in market interest rates were obtained for all assets considered. However, for two assets, the short- and intermediate-term assets, the standard errors were larger than the coefficients.

The model also suggested that as inflows of funds were expected, the proportion of cash held in the portfolio would decrease while the proportion of earning assets would increase. The regression results were again consistent with the hypothesis. The predicted directions were obtained for all assets. But again for two assets, in this case loans and the intermediate terms, the standard errors were larger than the coefficients.

Favorable results on balance were obtained in a separate test of the regression estimates with data drawn from a time period not used in the estimations. These results indicated that commercial banks' adjustments toward the equilibrium composition appear to be rather slow, taking approximately twenty months to reduce the cash share by about a third. Moreover, it seems that although the structure of interest rates appears to have small explanatory value considering short period changes, it nevertheless determines to a considerable extent the ultimate asset composition.



## DISCUSSION

MARTIN J. BAILEY: Pierce's paper raises problems of interpretation because it contains elements of the regression of a variable on itself. The denominator of the dependent variable  $L_t/F_{t-1}$  also appears as the denominator for two of the independent variables. That this denominator also appears as a variable in its own right helps to correct any bias arising from this source, but it follows that the coefficient of this term must be interpreted in a different way than Pierce did. The use of the discontinuity variable for upswings also involves an element of such self-regression, because the continuation of an upswing for still another month is in practice highly correlated with loan expansion, and in the *ex post* determination of turning points partly depends on it.

Malkiel's paper is a good restatement of the balance between expectations and portfolio preferences that determines the demand side of the term structure of interest rates. Regrettably, he slips into the conventional habit of referring to portfolio preferences as, in effect, the determinant of "Hick-sian liquidity premiums." It is true that the evidence, e.g., that presented by Kessel a year ago, overwhelmingly supports the proposition that short-term securities tend to have lower yields than long-term ones; that is, the term structure of the available supply of assets tends on the whole to run longer than the term structure that firms and the public would desire at a completely flat term structure. We should realize, however, that it is also possible for an asset to have a shorter maturity than its holder would desire at a flat yield curve. If the expected time at which he wants his capital in cash is later than the maturity of the asset, he gains no additional capital security but suffers some extra income uncertainty compared to that corresponding an asset of the appropriately longer maturity. The premium he requires to induce him to hold the shorter-term asset we might call a "solidity premium." That such solidity premiums sometimes exist is evidenced by the relatively low yields characteristic of Treasury 1995's.

Russell's paper raises the problems ordinarily associated with regressing a dependent variable on its own lagged values. In some cases the regression coefficient is painfully close to one, although its difference from one is significant in every case. If serially correlated errors are present in the true relationship or if serially correlated errors of measurement of the variables are present, these errors bias the other coefficients toward zero and the lagged dependent variable's coefficient toward one. Such errors may help to explain the very slow adjustment of portfolios that seems to be implied by Russell's results. They could also play a role in his apparently very good extrapolation results: The part of the extrapolation period in which the regression equations fit best is the part closest to the period for which the regressions were originally fitted.

ARTHUR OKUN: The three papers presented here deal with relationships among assets in financial markets. In a welcome development, most of the profession has adopted a general equilibrium view in its theorizing about the operations of financial markets. This view recognizes that a force

affecting one financial asset may have important repercussions on many others. It also highlights the fact that aggregate demand for goods and services may be affected by a wide variety of financial occurrences, including some which do not change monetary aggregates, such as bank reserves, money supply, or total liquid assets. But the empirical counterpart of this development has lagged behind: we still know unfortunately little about the substitution and complementarity relationships among various assets. The doctoral dissertations of Russell, Pierce, and Malkiel add to our knowledge and offer compelling evidence that these three men are likely to contribute further in these areas.

Russell and Pierce have examined the portfolio decisions of commercial banks. In both studies, it is clear that complex trade-offs affect the allocation of funds; it is heartening to see that interest rates come through loud and clear in their role as price allocators.

Both Russell and Pierce view banks as risk-averting units; the decision to diversify bank portfolios is explained by a willingness to sacrifice some return to get added safety. Otherwise, it would seem, banks might put all of their funds into the prospectively most profitable asset. In my opinion, the analysis of risk aversion and its implications for portfolio choice is one of the key developments of the last decade in monetary theory, but I am not happy about its use by Russell and Pierce. Risk aversion should influence the portfolio choices of individuals; it is not obviously applicable to corporations. If the bank's decisions are made on behalf of its shareholders, the desire for diversified personal portfolios is no argument for the bank managers to diversify. No stockholder will hold—certainly, none is compelled to hold—all of his assets in the shares of the bank; he is free to adjust the rest of his portfolio as he chooses. Just as Gordon and Gilbey are not obliged to put the vermouth in their gin bottles in order to serve a martini-loving world, the banker could, in principle, offer a straight equity to his shareholders, leaving the mixing to them.

I am not arguing that bankers are unaffected by risk aversion. There is a market for premixed martinis, and there may be reason to serve the consumer who wants bank stock mixed in a particular way to economize on the overall transactions costs of managing his entire portfolio. Further, the incentive system to bank management may create a rationale for risk aversion, even if diversification does not serve the stockholder especially well. But I do object to the simple and direct extrapolation of the portfolio balance model to bank behavior.

I feel free to criticize this framework because I have another to offer. The general pattern of banks' portfolios can be explained satisfactorily through the precautionary motive for liquidity. So long as deposit inflows and outflows are uncertain and so long as the costs of excessive and insufficient reserves are asymmetrical, even profit-maximizing banks will have good reasons to diversify their portfolios. The love of profit is enough to account for diversification. This approach is developed by my colleague, Richard C. Porter, in "A Model of Commercial Bank Portfolio Selection" (*Yale Economic Essays*, 1961).

•

•

Turning to the econometric work of the Pierce and Russell papers, I find it interesting that Russell's target adjustment estimates imply that bankers move very slowly toward their desired portfolios. The use of target adjustment models in econometric studies is very widespread and has a strong attraction; it seems very plausible that decision-makers know where they want to go but do not attempt to reach their destinations instantaneously. But nearly all the empirical results of such models imply that the movement toward the goal is surprisingly slow. I do not believe we live in a molasses world, as these persistent findings suggest. The determinants of the target are never specified with complete accuracy in our models and hence we measure the target imperfectly. Sometimes, an apparent off-target situation may reflect a revision of the target resulting from forces that are omitted from our equations. As a consequence, we may infer that the economic units are ambling along toward their destination, when in fact we are mistaken about their precise destination. In general, as we learn to specify the determinants of targets more accurately, I am convinced that we will find more rapid responses. In Russell's case, specification might be improved by considering the composition of bank liabilities (as among types of deposits and depositors) and the level of legal reserve requirements as determinants of asset choices of banks.

One of Pierce's variables puzzles me. The  $(L/F - \hat{L}/F)$  variable is meant to reflect two hypotheses: the banker is reluctant to expand the loan share of his portfolio far above "normal"; and he uses the volume of actual loans to assess the state of loan demand. The positive coefficient on this variable implies that the second influence is dominant. I suspect that the variable may also reflect lagged adjustments. On the whole, it seems to be serving too many masters to be a satisfactory proxy variable.

In analyzing the term structure of interest rates, Malkiel has unearthed important new material in a well-worked field and he deserves to be congratulated. In particular, he has displayed energy and ingenuity in his search for empirical evidence by reviewing bond letters, talking to portfolio managers, etc. Such evidence does not lend itself readily to quantification—much less to statistical testing. It is furthermore hard to report in a manner that is convincingly scientific. For these reasons, it is too likely to be ignored. It is gratifying that Malkiel was not afraid to dig into dirty data and that he emerged with clean hands.

The battle between the pure expectationists and the institutionalists had a clear issue: do relative supplies of long-term and short-term assets affect the term structure of interest rates? Once the obviously untenable assumptions of costless arbitrage and uniform expectations are modified in the expectationist model, the issue disappears. A modified expectationist should believe that relative supplies make a difference, as Malkiel demonstrates. Longs and shorts are not perfect substitutes, even though the world is not segmented. A priori arguments cannot determine the importance of relative supplies: it is a matter of degree which requires statistical analysis. My own empirical work for the Commission on Money and Credit suggested that yield differentials were quite insensitive to changes in relative supplies of

longs and shorts, but I consider the issue very much open and sorely in need of more econometric work, especially work pursuing a less aggregative approach.

Adopting a modified expectations approach to explain the term structure, one must resort to inelastic interest rate expectations to account for the much greater volatility of short rates than of long. The *a priori* case for inelastic expectations about bond prices is no stronger—indeed, no different—from the case for inelastic expectations about commodity prices. Yet, while we lean to the view that commodity prices have elastic expectations, we infer from the term structure data that interest rate expectations are inelastic. The modified expectationists—such as Malkiel and I—cannot be comfortable with that inference until it rests on stronger analytical grounds. A possible explanation may stem from the way capital gains (or losses) on existing long-term issues necessarily change their share of portfolios when rates change, or from the presence of lagged responses in the issuance of long-term bonds.

RICHARD E. QUANDT: The three papers presented have much in common, not only in that they all deal with important monetary phenomena, but they also share a substantial degree of excellence, as well as a certain similarity in the basic approach they employ. To varying degrees each paper attempts to construct a theoretical framework for the purpose of explaining some observable phenomena.

Because just so much can be accomplished in a limited time and since the topic of Mr. Malkiel's paper bears less resemblance to the other two than those two bear to each other, I shall concentrate my remarks on the papers of Mr. Pierce and Mr. Russell and specifically on some theoretical aspects. Since these papers are abridged versions of more extensive documents, my remarks are really questions to which the answers will undoubtedly be found in the unabridged versions.

Both papers are intimately concerned with the process of portfolio selection by commercial banks. Both posit commercial bank behavior to rest on utility maximization, and both attempt to build operational models into which real numbers can be plugged. In addition, their choice of a utility function is very similar; they both associate risk with the standard deviation of return and either posit or imply the existence of an efficiency frontier at which there is a trade off between expected return and risk. In other words, a greater risk, as measured by the standard deviation, can be tolerated if this greater risk is compensated by a higher expected return. In this connection two considerations emerge: (1) Baumol, in a recent paper, has pointed to some difficulties which may be inherent in such an approach if it is not possible to vary the expected return and the standard of return independently; (2) alternative measures of risk seem feasible, such as the worst possible outcome of a particular portfolio choice—which would be a very conservative measure indeed—or the probability of an outcome worse than some stated threshold, neither measure being necessarily monotonic in the standard deviation. Since the use of the standard deviation of return

•

as a measure of risk is a fairly natural consequence of the assumption of a quadratic utility function, one is led to ask whether other formulations could not be more satisfactory.

Another, and perhaps not completely unrelated point, emerges from Mr. Pierce's paper. He assumes that demand deposits have the lognormal distribution. Such an assumption is usually defended on the grounds that stochastic process governing the generation of deposits obeys the Law of Proportionate Effect; that is to say, that increments in individual deposits occur at random but are proportionate to the size of the deposit. But then, under suitable assumptions, the Law of Proportionate Effect can also lead to the Pareto distribution which, for an  $\alpha$ -coefficient of less than 2 does not have finite variance! If we "fix" this difficulty by assuming—in the absence of concrete tests—that the Pareto  $\alpha$  is greater than 2, one can still demonstrate that under the remaining assumptions of Mr. Pierce, the bank considers the standard deviation of deposits in period  $t+1$  to be proportional to the current level of deposits. Slightly different assumptions about deposits and the creation of new deposits lead to still different distributions such as the Yule distribution. One wonders, therefore, as to the precise significance of introducing the lognormal distribution at this point; it does not seem to play an important role in his arguments. It seems, however, that this does not detract from Mr. Pierce's imaginative handling of how to formulate the estimating equation and, in particular, his apt use of a dummy variable to allow his indicator of the desirability of the bank's portfolio position to have differential effects in times of expansions and contractions of loan demand. There may be some identification problem with respect to loan demand since we have only a measure of loans extended, but the empirical results seem convincing. Finally, one may note that the arbitrary constant added to  $(L/F - \hat{L}/F)$  in order to enable us to take the logarithm of that quantity could have been estimated itself—thus making the computation problem much more difficult but the estimates more efficient.

Mr. Russell pays special attention to the dynamics of portfolio adjustment and the attempt to account for the dynamics is creditable in itself. He accomplishes this by specifying a Markov process in which the states are abstract asset classes. He calculates the approximate transition probabilities as functions of the inherent rate differentials and the equilibrium path of the system. One may question the use of a one-period Markov process which implies a very short memory on the part of the system and one wonders whether alternative formulations of the dynamics—a model with a longer memory such as a distributed lag model—might not have yielded even better empirical results.

# INVITED DISSERTATIONS, II

## THE DETERMINANTS OF VOLUME IN A PROBABILISTIC MARKET\*

By FREDERICK T. SPARROW  
*Johns Hopkins University*

The purpose of this paper is to present an intuitive model for determining the expected volume, or quantity traded, in a market that is known to contain substantial random elements, and to test some hypotheses about the determinants of expected volume suggested by the intuitive model against the results of an exact, but extremely un-intuitive, formulation of expected volume described in the mathematical appendix of this paper.

In another paper,<sup>1</sup> I have argued that if we think of the traditional demand and supply curves and the change of price function as lines connecting the means of a series of Poisson distributions, as in Figures 1 and 2, such a market can be represented as a Markov process whose states are all possible combinations of price and excess demand or supply. The Markov process so constructed can be solved for the steady-state probability distribution of price and excess demand or supply, from which one can derive the mean and variance of the marginal probability distribution of price. A logical extension of this work is an investigation of the mean volume of such a market, referred to hereafter as  $E(V)$ , reconstituted to include random elements.

If we define  $E(V)$  to equal one over the expected time between sales (thus, if the expected time between sales was one-thirtieth of an hour, we would expect the volume to be thirty an hour), we can obtain  $E(V)$

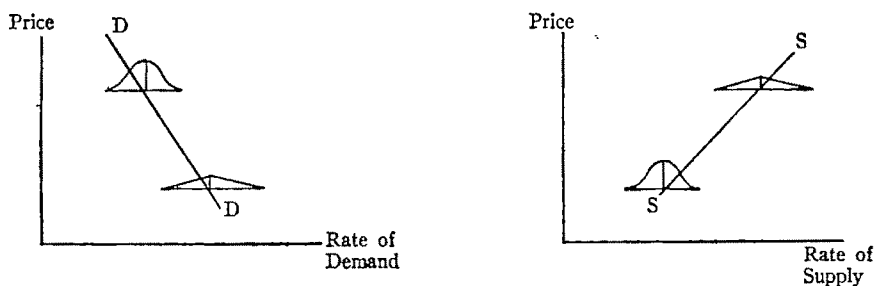


FIGURE 1  
PROBABILISTIC DEMAND AND SUPPLY

\* I would like to thank Takeshi Amemiya, Carl Christ, R. V. Evans, and Edwin Mills for their many helpful suggestions during the preparation of this paper.

<sup>1</sup> "Some Experiments with a Queueing Theory Model of Market Equilibrium," forthcoming.

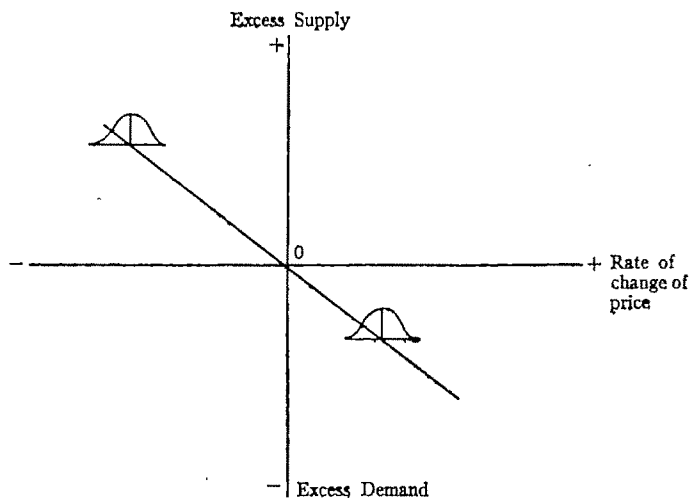


FIGURE 2  
PROBABILISTIC CHANGE OF PRICE

by calculating the expected time between sales. The derivation of the expected time between sales is given in the mathematical appendix. A quick glance at the appendix will show that this model adds little to our intuition about the determinants of expected volume: it will obtain the correct answers, but the factors that govern the answers are obscured by the rather complicated mathematics.

Since our goal is to give insights into the factors that govern the expected volume rather than calculate exact answers, let us attempt to construct a "naïve" model of expected volume that does give such insights at the expense of mathematical exactness and check the results of the intuitive model against the exact formulation's conclusions.

### *The Intuitive Model*

Consider Figure 3, a typical plot of demand and supply in a market.  $P_D$ , the deterministic equilibrium price, and  $V_D$ , the deterministic equilibrium volume, are determined by the intersection of the demand and supply constructs. Suppose we had at our disposal only the following information about a market that contained substantial random elements:

1. The *expected* rate of arrival of customers for each price  $j$ ;
2. The *expected* rate of arrival of goods for each price  $j$ ;
3. The steady-state probability that there would exist excess demand for each price  $j$ ;
4. The steady-state probability that there would exist excess supply for each price  $j$ ;

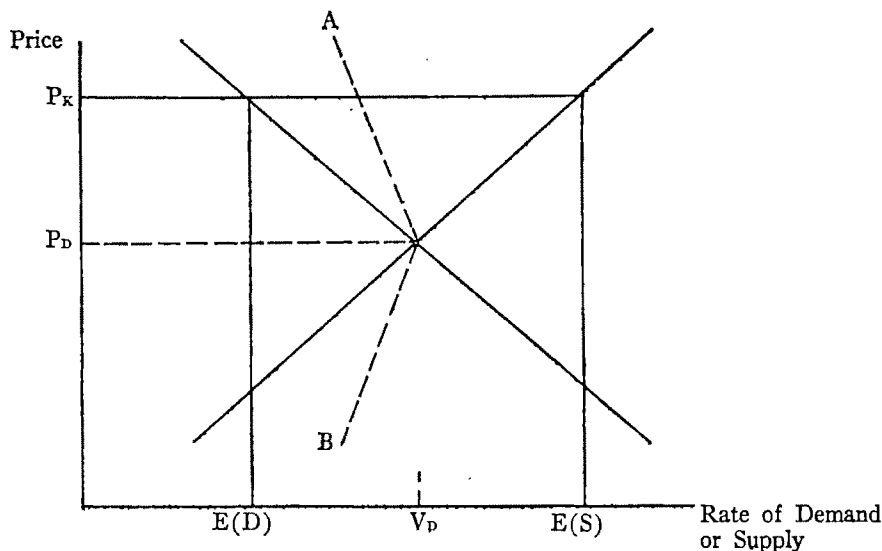


FIGURE 3

5. The steady-state probability distribution for the price  $j$ ;
6. The steady-state probability that there exists neither excess supply nor excess demand for each price  $j$  is negligible.

What can be said about the expected volume in such a market, if the price is  $P_K$  in Figure 3? If there is excess supply associated with price  $P_K$  at a given moment, then it is the mean arrival rate of people,  $E(D)$  in Figure 3, that determines the quantity traded. Conversely, if there is excess demand, the mean arrival rate of goods,  $E(S)$  in Figure 3, controls the volume. Then the expected volume on condition the price is  $j$  can be expressed as:

$$E(V | p = j)$$

$$(1) \quad = Pr \{ \text{excess demand} | p = j \} \cdot E(\text{arrival rate of goods} | p = j) \\ + Pr \{ \text{excess supply} | p = j \} \cdot E(\text{arrival rate of customers} | p = j)$$

Let the dotted line  $AB$  in Figure 3 be the plot of each of these conditional volumes, the line necessarily lying between the demand and supply curves. Now, if we utilize our knowledge about the steady-state distribution of the price  $j$ , we may calculate the unconditional expected volume of such a market by weighting each of the conditional volumes by the probability that price  $j$  will occur, and sum over all terms. Thus,

$$(2) \quad E(V) = \sum_{j=0}^n E(V | p = j) \cdot Pr\{p = j\}, \quad n = \text{upper limit on price.}$$



Apparently, considering equations (1) and (2), there are three factors that influence  $E(V)$ : (a) the expected arrival rates of customers and goods, i.e., the demand and supply curves, (b) the probabilities that excess demand or supply are associated with each price, i.e., the weights to be attached to the expected arrival rates in order to construct the points on the  $AB$  curve, (c) the steady-state distribution of price.

### *The Hypotheses*

What relations between  $E(V)$  and the economic variables (the probabilistic supply, demand, and change of price functions) are suggested by the above argument?

Consider factor *a* first. If the means of the probabilistic demand and supply curves increase, the expected volume increases. This is equivalent to the conclusion in a deterministic world that volume increases when the demand and/or supply curves shift to the right. But note that in a probabilistic market, even if the intersection of the mean demand and supply curves remain constant, any clockwise rotation of the mean demand or supply curves above their intersection or any counterclockwise rotation below the intersection will increase the expected volume of the market.

Thus, in Figure 4, even though  $\{D_1, S_1\}$  and  $\{D_2, S_2\}$  have the same deterministic volume, we would expect a higher  $E(V)$  to be associated with  $\{D_2, S_2\}$  than  $\{D_1, S_1\}$ . This hypothesis was tested by a series of

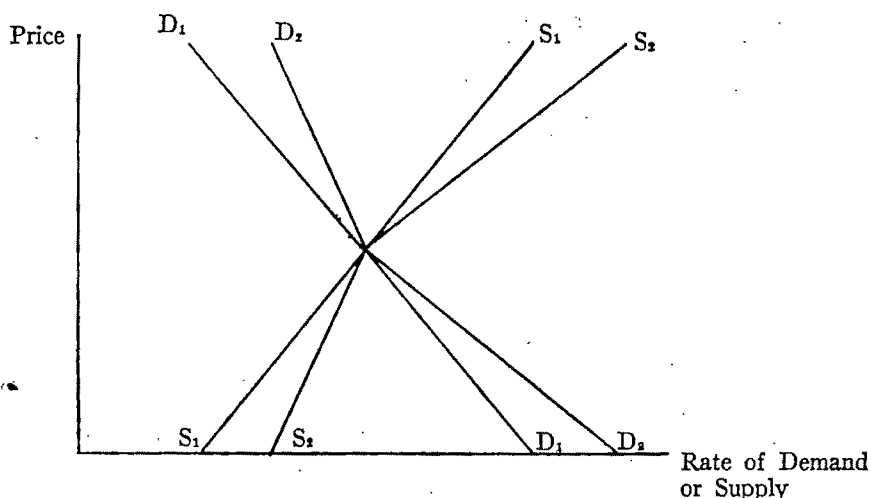


FIGURE 4

experiments where the exact  $E(V)$  was calculated for markets with given probabilistic demand and supply curves by the method described in the appendix. The above conclusion was supported in all experiments.

In constructing hypotheses based on factors  $b$  and  $c$ , we are faced with a dual question: first, what economic variables act on  $E(V)$  via factors  $b$  and  $c$ ; second, what do  $b$  and  $c$  suggest about the nature and direction of the relation between  $E(V)$  and the economic variables?

Experimentation has shown that although the demand, supply, and price-change distributions jointly determine  $b$  and  $c$ , it is the price-change function that primarily governs their values. To answer the second question, let us first examine the effect of the price-change function on  $b$ , the conditional probability of excess demand or supply for a given price. First, the experiments have shown that when the price-change mechanism is "weak" (i.e., large stocks of excess demand or excess supply produce small changes in price) high prices are associated with stocks of excess supply, and low prices with stocks of excess demand, an intuitively appealing result. That is, the steady-state probability distribution is concentrated along the diagonal of the state space plane, running from the high price-excess supply corner to the low price-excess demand corner; this concentration grows as the price-change mechanism grows weaker, and presumably in the limit all the probability lies on this diagonal. Conversely, the experiments show that strong price-change mechanisms result in high prices associated with stocks of excess demand, and low prices associated with stocks of excess supply; i.e., the steady-state probability distribution tends to concentrate on the diagonal running from the high price-excess demand corner to the low-price-excess supply corner. Since this result does some violence to the intuition, perhaps an attempt to explain this phenomena is in order. If the price-change mechanism is strong, then as soon as any excess supply (demand) accumulates, prices are driven down (up) very rapidly; thus for the bulk of the duration of the accumulation of supply (demand) prices are low (high). Hence the result that excess demand is associated with high prices, and excess supply with low prices.

What are the implications of this on the conditional probabilities of  $b$ , and hence on the shape of the  $AB$  curve? Figures 5a to 5c illustrate the shape of the  $AB$  curve for very weak, moderate, and very strong price-change mechanisms, respectively. For the very weak case, the  $AB$  curve approximates the demand curve above equilibrium, since it is almost certain that high prices will be associated with excess supply; hence the rate of arrival of demanders primarily governs volume. Similarly, the  $AB$  curve approximates the supply curve below equilibrium, since it is almost certain that stocks of excess demand will be associated with low prices. Note that the conventional conclusion that the demand curve governs volume when prices are above equilibrium and the supply curve governs volume for prices below equilibrium is true only in the limiting case of an infinitely weak price-change mechanism. Figures 5b and 5c can be supported by similar arguments.

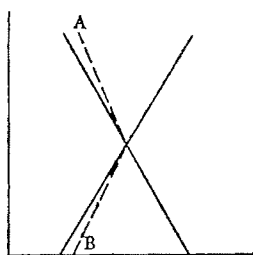


FIGURE 5a

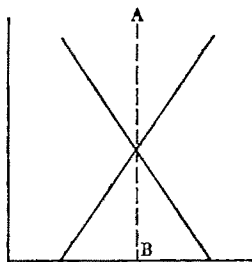


FIGURE 5b

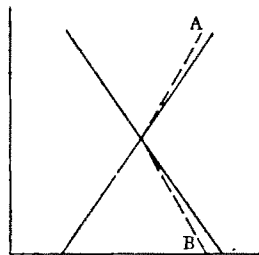


FIGURE 5c

What can one infer about the influence of the price-change mechanism on  $E(V)$  from the above argument? One would expect that the stronger the price-change mechanism, the higher the expected volume—all else equal. Unfortunately, all else is not equal, since the price-change function strongly affects factor  $c$ , the steady-state probability distribution of price. Hence we must withhold judgment on the ultimate effect of the price-change mechanism on  $E(V)$  until we have investigated its relation with the price distribution.

The effect on the price-change distribution revealed by experimentation follows a pattern well known to systems analysts; as the strength of the price-change mechanism increases, the variance of the price distribution first decreases, reaches a minimum, and then rapidly increases.<sup>2</sup> What we are observing is the effect of the strength of a feedback mechanism on the controlled variable; for a while, any increase in the feedback strength will make the controlled variable more stable—i.e., it will reduce its variance; but after a point, the feedback mechanism becomes so strong that it tends to “over correct,” and the system becomes unstable—i.e., the variance of price increases.

How does the price-change function affect  $E(V)$  through its influence on the price distribution? If the  $AB$  curve were frozen in the position pictured in Figure 5a then as the price-change mechanism grew in strength,  $E(V)$  would first increase, and then decrease. Conversely, if the  $AB$  curve were as in Figure 5c,  $E(V)$  would decrease and then increase. Since the price variance and the  $AB$  curve change simultaneously as the strength of the price-change mechanism grows, there is no conclusive a priori argument about the net effect of the price-change mechanism on  $E(V)$ .

The experiments, on the other hand, are unequivocal in their outcome; using “reasonable” assumptions about the shape of the demand

<sup>2</sup> We assume that the price-change mechanism does not affect  $E(P)$ . A necessary condition for this to occur is that the price mechanism be “symmetrical”; i.e., equal stocks of excess demand or supply produce price changes of equal absolute value. For a fuller discussion, see the reference in footnote 1.

and supply curves,  $E(V)$  increased as the strength of the price-change mechanism increased.<sup>3</sup>

### Summary

What has this treatment added to the fund of knowledge about volume, or quantity traded, in the market place? First, it presents two models, one intuitive and one exact, that are useful in developing and testing hypotheses about the influence of the economic variables on  $E(V)$ . Briefly stated, these conclusions are: (1) An increase (decrease) in the mean rate of demand or supply for any price will produce an increase (decrease) in  $E(V)$ . (2) In general, an increase (decrease) in the strength of the price-change mechanism results in an increase (decrease) in  $E(V)$ . More importantly, the treatment suggests that the conventional conclusion that volume is governed exclusively by the demand curve above and the supply curve below the equilibrium price holds true only in the unlikely case of an infinitely weak price-change mechanism. The models show that except in limiting cases, both the demand and supply curves enter into the determination of  $E(V)$  at all prices.

### MATHEMATICAL APPENDIX\*

First, we need to define the concept of a sales event. If we are in some particular state  $(i, j)$  of the system at time  $t$ ,  $i$  indexing the magnitude of the excess demand  $[i-]$  or supply  $[i+]$  and  $j$  the price, three events may take place in a small interval of time  $\Delta t$  following  $t$ :

- A. A good can arrive with probability  $\lambda_j \Delta t$ , if, as is assumed the time between arrivals of goods on condition the price is  $j$ , is negative exponential with mean  $\lambda_j$ ;
- B. A person demanding a good can arrive, with probability  $\mu_j \Delta t$  under similar assumptions about the distribution of time between arrivals;
- C. A price change can take place with probability  $\alpha_i \Delta t$ , if the time between price changes on condition excess demand or supply is  $i$ , is negative exponential with mean  $\alpha_i$ .

A sales event is an event which results in a person and a good leaving the market; all other events will be called regular events. Thus, if there exists a queue of goods (excess supply), a good arrival event, which represents a transition from  $(i, j)$  to  $(i+1, j)$ , and a price change event, which represents a transition from  $(i, j)$  to  $(i, j-1)$ , are regular events, while the arrival of a demander which moves the system from  $(i, j)$  to  $(i-1, j)$  is a sales event. If there is excess demand at time  $t$  (people or orders to buy waiting for goods) only the good arrival event, which moves the system from  $(i, j)$  to  $(i+1, j)$ , is a sales event.

\* This, of course, does not rule out other outcomes. In all probability other results can be obtained, but only by making assumptions that are pathological in nature.

\* These results were first presented by R. V. Evans, "Time Interval Random Variables for a Class of Two Dimensional Queues," Western Management Science Institute Working Paper #43, U.C.L.A. Graduate School of Business Administration, Oct., 1963.

Our task is to determine the expected time between such sales events, referred to as  $E(TBS)$ , for the system as a whole. First, define:

- (1)  $\Omega$  = the set of states, and
- (2)  $\Omega'$  = the subset of states where a sales interval may start. Let  $\omega$  be a state in  $\Omega'$ .

We wish to calculate  $Pr$  {system is in state  $\omega$  at time  $t$  |  $\alpha$  sales event has just occurred}, referred to as  $Pr\{\omega|S\}$ , and  $E$  (time until the next sales event | in state  $\omega$ ), referred to as  $E(TUNS|\omega)$  in order to calculate;

$$(3) \quad E(TBS) = \sum_{\omega \in \Omega'} E(TUNS|\omega) \cdot Pr\{\omega|S\}$$

First, consider  $Pr\{\omega|S\}$ . We need to define

- (4)  $\Omega''$  = the sub set of states from which it is possible to leave via a sales event, and
- (5)  $\Gamma$  = a subset of  $\Omega''$  from which it is possible to reach  $\omega$  via a sales event.

Let  $\gamma$  be a state in  $\Gamma$ ,  $P(\gamma)$  be the steady-state probability of being in state  $\gamma$  at time  $t - \Delta t$  [derived from the solution of the Markov process mentioned in the body of the paper], and let  $P(\gamma)$  be the probability of a sales event transition in  $\Delta t$  from state  $\gamma$  to state  $\omega$ . Then

$$(6) \quad Pr\{\omega|S\} = \sum_{\gamma \in \Gamma} P(\gamma)\alpha(\gamma) / \sum_{\omega \in \Omega'} \sum_{\gamma \in \Gamma} P(\gamma)P(\gamma)$$

To compute  $E(TUNS|\omega)$ , let the event  $A$  be the occurrence of a sales event in a time period  $t$  on condition we start in state  $\omega$ . Then  $A$  is the union of all possible sequences of disjoint events that start in  $\omega$  and end in a sales event, which contain no other sales events except the terminating event. Let  $H_i$  be such a sequence  $i$  events long. Then

$$(7) \quad Pr\{A\} = Pr\{A \cap H_1\} + Pr\{A \cap H_2\} + \dots + Pr\{A \cap H_n\},$$

$n$  = upper limit on the length of the sequence.

(N.B. There may be more than one sequence of length  $H_i$ .)

Then

$$(8) \quad E(t) = E(TUNS|\omega) = E(t|H_1) + E(t|H_2) + \dots + E(t|H_n)$$

Consider  $E(t|H_1)$ , and assume  $\omega$  is a state where excess supply exists; i.e.,  $\alpha$  and  $\lambda$  are the mean time between arrivals of regular events and  $\mu$  the mean time between arrivals of a sales event, assuming, as before, all follow the negative exponential distribution. Then

$$(9) \quad \begin{aligned} Pr\{t|H_1\} &= Pr\{\alpha \text{ and } \lambda \text{ events } \geq t\} \cdot Pr\{\mu \text{ event at } t\} \\ &= e^{-(\lambda+\alpha)t} \mu e^{-\mu t} dt = \mu e^{-(\lambda+\mu+\alpha)t} dt \end{aligned}$$

Multiplying (9) by  $[\lambda + \mu + \alpha / \lambda + \mu + \alpha]$  we obtain

$$(10) \quad Pr\{t | H_1\} = \left( \frac{\mu}{\lambda + \mu + \alpha} \right) (\lambda + \mu + \alpha) e^{-(\lambda + \mu + \alpha)t} dt$$

which is  $(\mu/\lambda + \mu + \alpha)$  times the probability one stays in  $\omega$  units of time. Then

$$(11) \quad E(t | H_1) = \left( \frac{\mu}{\lambda + \mu + \alpha} \right) \left( \frac{1}{\lambda + \alpha + \mu} \right)$$

which is the probability the sequence occurs times the expected time spent in  $\omega$ .

Consider now  $E(t | H_2)$ . Suppose the regular event  $\alpha$  leads us to a state  $\zeta$  which in turn can be left by a sales event with mean interval time  $\mu^*$ , or other regular events with mean interval times  $\lambda^*$  and  $\alpha^*$ . Then

$$(12) \quad Pr\{t | H_2\} = \int_{\tau=0}^t e^{-(\lambda + \mu)\tau} \alpha e^{-\alpha\tau} e^{-(\lambda^* + \mu^*)(t-\tau)} \mu^* e^{-\mu^*(t-\tau)} d\tau dt$$

Multiplying top and bottom as before,

$$(13) \quad Pr\{t | H_2\} = \int_{\tau=0}^t \left( \frac{\alpha}{\lambda + \mu + \alpha} \right) (\lambda + \mu + \alpha) e^{-(\lambda + \mu + \alpha)\tau} \cdot \left( \frac{\mu^*}{\lambda^* + \mu^* + \alpha^*} \right) (\lambda^* + \mu^* + \alpha^*) e^{-(\lambda^* + \mu^* + \alpha^*)(t-\tau)} d\tau dt$$

or

$$(14) \quad Pr\{t | H_2\} = \left( \frac{\alpha}{\lambda + \mu + \alpha} \right) \left( \frac{\mu^*}{\lambda^* + \mu^* + \alpha^*} \right)$$

· convolution of the probabilities of staying in  $\omega$  and  $\zeta$  during  $t$ .

Then

$$(15) \quad E(t | H_2) = \left( \frac{\alpha}{\lambda + \mu + \alpha} \right) \left( \frac{\mu^*}{\lambda^* + \mu^* + \alpha^*} \right) \left( \frac{1}{\lambda + \mu + \alpha} + \frac{1}{\lambda^* + \mu^* + \alpha^*} \right)$$

or

$$(16) \quad E(t | H_2) = Pr \{ \text{sequence } H_2 \text{ occurs} \} \cdot (\text{sum of expected time spent in } \omega \text{ and } \zeta). \text{ In general,}$$

$$(17) \quad E(t | H_n) = Pr \{ \text{sequence } H_n \text{ occurs} \} \cdot (\text{sum of the expected times spent in all states in the sequence } H_n), \text{ and}$$

$$E(t) = E(TUNS | \omega) \text{ may be formed by summing all these terms.}$$

Substituting (6) and (8) into (3) and summing over all  $\omega$  gives the required result, the expected time until the next sale.

## THE FEDERAL RIVERS AND HARBORS PROGRAM: AN ANALYSIS OF REGIONAL IMPACT\*

By ROBERT HAVEMAN  
*Grinnell College*

In recent years, federal water resource investment appropriations to the Civil Works program of the U.S. Army Corps of Engineers have averaged nearly a billion dollars annually. Inasmuch as this entire appropriation is a politically allocated domestic expenditure, it is clear that the sectional impact of such federalism is a concept worthy of analysis. Moreover, because of the increasing use of regional income differentials as a criterion for the allocation of federal appropriation, it is essential that those factors which determine the regional impact of any federal expenditure program be well-defined and clearly understood.

Basically, four alternative approaches have been suggested for evaluating the regional impact of development programs. They are: (1) use of an aggregate economic indicator, such as personal income, to compare relative changes in a region with a development program with relative changes for other regions, (2) detailed case-by-case analyses of the effects of an expenditure "on the parameters of the region's production and consumption functions over time and the trend in the structure of markets,"<sup>1</sup> (3) use of regional interindustry models in tracing the impact of a government expenditure program on the various output sectors of a region, and (4) comparison of the direct estimates of costs and expected returns (in terms of real income produced) which accrue to a region from an expenditure in an attempt to determine the net impact of the expenditure on the region. In various studies each of the first three of these approaches has been utilized with varying degrees of success.<sup>2</sup> In this paper, the final approach is used in an attempt to secure a more satisfactory evaluation of the impact of a federal program, namely, the Corps program, on a particular section of the country during a particular period of time.

\* The author wishes to thank Professor Rendigs Fels, of Vanderbilt University, for his most helpful comments and criticisms of the various drafts of the doctoral dissertation from which this paper is drawn. Also, thanks are due my colleagues at Grinnell College for their discussions and comments and to the Inter-University Committee for Economic Research on the South for its financial support.

<sup>1</sup> John V. Krutilla, "Criteria for Evaluating Regional Development Programs," *A.E.R.*, May, 1955, p. 132.

<sup>2</sup> See, for example, B. U. Ratchford, "Government Action or Private Enterprise in River Basin Development: An Economist's View," *A.E.R.*, May, 1951, pp. 299-307, Krutilla, *op. cit.*, pp. 120-32 and R. A. Kavesh and J. B. Jones, "Differential Regional Impacts of Federal Expenditures: An Application of the Input-Output Matrix to Federal Fiscal Policy," *Regional Sci. Assn. Proceedings*, II, 1956, pp. 152-67.

The region to be considered is composed of ten southern states<sup>3</sup> and the period of time which the study covers is the seventeen years from 1946 through 1962. First, a brief statement concerning the absolute level of appropriations to the region will be presented and, second, a more technical analysis of the net impact of the program on each state will be discussed. This latter analysis will take into account both the costs and the returns, primarily in the form of flood control and navigation benefits, which are expected to accrue to the state because of the Corps program.

## I

During the seventeen-year period covered by the study, the entire Corps program received approximately \$8 billion in federal appropriations.<sup>4</sup> Of this total, more than one-fourth was allocated to the ten southern states.<sup>5</sup> Thus, while the national per capita appropriation was about \$50, the southern appropriation was \$72 per person.<sup>6</sup> On a per capita basis, then, the South appears to have received nearly one and one-half times the appropriation of the country taken as a whole and nearly one and two-thirds times the per person appropriation as the non-South.

Moreover, not only cross-sectionally has the South fared rather well, but also over time. During the seventeen-year period, while the average annual increase in total Corps construction appropriations was less than \$42 million, the average annual increase in Corps appropriations devoted to the South was over \$11 billion. Thus, whereas the average annual percentage rate of growth of the total Corps program during the period was 13.7 percent, the comparable figure for the appropriation dedicated to the South was 16 percent.

It is, therefore, clear that the South has been favored in the receipt of such water resource development funds both cross-sectionally and over

<sup>3</sup> The ten states are Alabama, Arkansas, Florida, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, South Carolina, and Tennessee. In dealing with problems of regional impact, this grouping is an appropriate choice, it constituting the lowest income region of the country. In per capita income terms, these states ranked, respectively, 47, 49, 30, 42, 46, 43, 50, 44, 48, and 45. With the exception of Florida they are the lowest nine out of the fifty states.

<sup>4</sup> The total Corps program includes two separate budgets, the General Construction budget and the Mississippi River and Tributaries Construction budget. The General Construction budget is by far the most substantial of the two, comprising about 90 percent of the total construction budget.

<sup>5</sup> The data on total appropriations was computed from information in the House Reports on the Senate-House Conference on each of the annual Civil Functions Appropriation Bills from 1946 through 1962 and from other miscellaneous bills containing Corps appropriations. In each case the appropriation was allocated to the state in which the project was located. In the cases in which a project was in two or more states, the appropriation was divided equally among the states. In the cases in which the project was a comprehensive river basin project flowing through two or more states, the appropriation was allocated to the state according to the actual construction expenditure which occurred in that state.

<sup>6</sup> The state population data used for the per capita estimates are averages of the 1950 and 1960 census figures for each state.



time. Not only are southern appropriations a rather large proportion of the total pie, but also a growing proportion.

## II

Given the implications of this preliminary data, the remainder of the paper will attempt to verify the existence of the favorable redistributive impact on the ten southern states resulting from the Corps program and, through a more refined approach, to estimate the magnitude of the impact by state and by time period. The method used in the study combines the concept of an income redistribution function with direct estimates of the "returns" and costs to a state's residents due to the Corps General Construction program.<sup>7</sup>

*Method and Assumptions.* By first presenting a general income redistribution function and then refining it to fit the situation at hand, an estimation of the real exchange of income can be made. Such a function for any given region,  $x$ , would be:

$$I_x = R_x - C_x \quad (1)$$

in which  $I_x$  is the net income redistribution to region  $x$  from a given federal program,  $R_x$  is the gross return derived by the residents of this region from the program, and  $C_x$  is the actual cost levied upon this group in order to obtain the return. This concept of income redistribution, it will be noticed, deviates somewhat from the commonly accepted concept. Whereas the common meaning connotes the existence of a fixed stock of whatever is being redistributed (so that a loss to one group necessitates an equivalent gain to another), the present concept does not. Moreover, the concept as used here deals with only one of the two or more groups involved. That is, if a group enters a transaction and if, from the transaction, it received in value more than it is forced to give up, this method would lead to the conclusion that income is redistributed toward that group.

If it is tentatively assumed that both the returns from and charges for the program occur during a period of time in which the group involved is indifferent to the time pattern of returns and charges, say a single moment of time, the determination of the variables in the function would be greatly facilitated. In such a case the returns would be equal to the maximum amount which the recipients of the outputs would be willing to pay in order to receive them (i.e., the area under the aggregate demand curve for the outputs) and the charges would be equal to the out-of-pocket costs to the group in that moment of time.

In the absence of such an abstraction, the simplified world of purity

<sup>7</sup> It should be noted that the analysis presented in this section excludes activity under the Corps Mississippi River budget.

vanishes and the world of the best estimate comes to take its place. Presuming then that such indifference to time patterns does not exist, it is essential that the earlier function (1) be modified so as to incorporate an allowance for time preference. One such feasible modification would be to measure both the group's willingness to pay for program outputs and the necessary charges required from the group to secure these returns as the group would view them at a given moment in time. In this case, the function would become

$$I_{x,t} = P(R_x)_t - P(C_x)_t. \quad (2)$$

Thus, the net redistribution of income to group  $x$  attributable to a given program as evaluated at time  $t$  ( $I_{x,t}$ ) equals the present value of the stream of gross benefits to the group evaluated at time  $t$  [ $P(R_x)_t$ ] less the present value of the stream of costs borne by the group also evaluated at time  $t$  [ $P(C_x)_t$ ].

Accepting this general approach to the empirical estimation of regional impact, it becomes essential that both the several measurement problems encountered and the approximations used in the remainder of the study be explicitly stated and analyzed. First, consider the income supplementing term of the equation [ $P(R_x)_t$ ]. For an empirical purist, such a measurement of present real value poses an impossible task for the discipline in its present state of development. This is so because the concept of real value can only be approached through an evaluation of the willingness to pay of the recipients of the outputs including the empirically unmeasurable concept of consumers' surplus. Moreover, as soon as the concept "value" is mentioned, the existence of a price is implied, and if this "price" is to be a meaningful concept in the analysis, it must be a perfectly competitive price determined in the market by producers who maximize nothing but profit and consumers who maximize only utility and which price equals the marginal cost of production. Also, the measurement of psychic gain through the tool of an aggregate demand curve necessitates constancy in the marginal utility of income to consumers. Still further, because the estimate is to be made at a given point in time, an appropriate interest rate must be used to discount the expected future stream of benefits. In order to accurately estimate the real effect, this rate must reflect the social rate of time preference of the beneficiaries of the project. Finally, the effects of the program outputs, both direct and indirect, must be confined to the region within which the project is located.

Realizing the extreme problems which these conditions pose for but a single project to say nothing of an entire program, several reasonable approximations must be made to insure practical measurability. First, it is assumed that the present value of the expected stream of outputs

at the time of project construction approximates the willingness to pay for these outputs by the recipients of them. This present value figure is computed from the estimate of annual project benefits as used by the Corps in their benefit-cost analyses and was secured from the twelve southern district Corps offices for each of the 165 projects used in the study. This estimate is then allocated to the state in which the project is located during the period of project construction (the proportion of total present value allocated in any given year being determined by the fraction of total federal construction cost which was appropriated to the project in that year). The pitfalls which accompany the use of this data are well known and recognized, but, because no other more accurate yet feasible alternative is in sight, these Corps data are accepted as an adequate approximation of the real value of project outputs.<sup>8</sup> Second, because of the need to discount an estimated stream of expected future project outputs in order to secure an approximation of real present value, an appropriate rate of time preference must be obtained. Further, because of the impossibility of obtaining the time preference pattern of the beneficiary group of every project, this must likewise be a reasonable surrogate. Consequently, a rate of 4.5 percent is accepted as an estimate of the national rate of time preference and is used to approximate the regional rate of time preference.<sup>9</sup> Third, it is assumed that the economy is operating at full employment, so that the initial expenditure merely diverts resources.<sup>10</sup> Finally, it is assumed that the benefits yielded by the projects are confined to the residents of the region within which the project is located and, moreover, that these additional expected benefits yield no secondary multiplier effects.<sup>11</sup>

<sup>8</sup> For a discussion of estimation techniques and their limitations, see Otto Eckstein, *Water-Resource Development: The Economics of Project Evaluation* (Harvard Univ. Press, 1958), pp. 110-258.

<sup>9</sup> William L. Miller, "The Magnitude of the Discount Rate for Government Projects," *So. Econ. J.*, April, 1962, pp. 348-56. Miller's study is a modification of a study by Otto Eckstein in which he estimated the proper rate to be 5.5 percent for the year 1955. See Otto Eckstein and John Krutilla, *Multiple Purpose River Development: Studies in Applied Economics* (Johns Hopkins Press, 1958), pp. 78-133.

<sup>10</sup> Alternatively, it could be assumed that the initial income effect, inasmuch as it does more than merely divert resources, would, because of the wide geographical dispersion of the expenditure, accrue entirely to residents outside the region.

<sup>11</sup> In performing the actual calculation, a minimum and a maximum estimate were made for the present value of estimated project benefits in order to make some allowance for price level changes during the period of construction. The higher of the two estimates is based on the assumption that the relationship between inputs and outputs is a physical relationship correctly estimated by Corps and expressed in monetary terms. Any deviation of actual from estimated construction cost is therefore assumed to be a simple price level change affecting both inputs and outputs uniformly. The Corps claims that, during a postwar period which they analyzed, 57.8 percent of such cost differences were due to such price level changes. See U.S. Army Corps of Engineers, *Report on the Federal Civil Works Program as Administered by the Corps of Engineers*, 1952, pp. 93-102. The lower of the two estimates is based on the assumption that all deviations of the actual construction cost from the expected construction cost were a result of only construction material price changes. Inasmuch as actual deviations are due to a combination of both of these factors, the correct estimate must lie within the range dictated by this maximum and minimum. Thus, in the statistical results presented later in the study, the benefit estimates are medians of the maximum and minimum estimates.

In evaluating the necessary charges to the region, that is, the income depleting term of the function  $[P(C_s)_t]$ , the method of estimation and the assumptions upon which it is based must again be made explicit. The necessary charges which are borne by southern states attributable to the postwar Corps program appear to fall into two mutually exclusive categories: those charges borne directly and those borne indirectly. Of those charges borne directly, two types can be distinguished: (1) the initial out-of-pocket charges borne locally and (2) those future operation costs for which local interests assume responsibility. The indirect charges are somewhat more nebulous, but again two types can be distinguished: (1) those federal taxes borne by each state in support of total Corps General Construction appropriations and (2) the state's share of future estimated operation costs on all Corps projects for which the federal government assumes responsibility.<sup>12</sup> As with the income supplementing term of the equation, the empirical estimation of these variables is most problematical. Questions of real value, interest rates, and allocation of burden again pose insurmountable problems to the empirical purist. Again, reliance must be placed on manageable, yet reasonably accurate, approximations.

The method used in the study to estimate the direct costs to a state's residents is fairly straightforward. For both the initial out-of-pocket costs and the locally-borne future operation costs, the Corps of Engineers estimates at the time of project construction are used. The present value figure for the future charges is again obtained by using the 4.5 percent rate of interest. As with the regional benefit estimate, the proportion of the total present value of a project's direct costs which are allocated to the region in any given year is determined by the fraction of total federal construction cost which was appropriated to the project in that year.

Estimation of the indirect regional charges requires a rather more complicated technique. In order to estimate those federal taxes borne by the residents of any given state in support of each year's federal appropriation to the Corps General Construction program, a modifica-

<sup>12</sup> There is one additional cost which falls into neither of these categories; namely, those costs which are payments for project benefits. Of the total annual benefits which resulted from the construction of the 165 southern projects, about 20 percent were of this reimbursable nature. However, because the reimbursements accrue to the general fund of the U.S. Treasury, they cannot be allocated to any one function or group but rather belong to the entire body of taxpayers. Because the national expected gain from these payments equals the national expected loss, the net gain (or loss) to any given region within the nation depends upon the relationship of the relative incidence of total expected reimbursable costs to the relative incidence of federal taxes (assuming that such reimbursable collections will, because they accrue to the general fund of the Treasury, reduce federal taxes in proportion to regional federal tax incidence). By assuming that the regional incidence of the reimbursable costs of Corps projects is identical to the regional incidence of federal taxes, such payments were excluded from the study. Consequently, because of the relatively small federal tax load borne by the South combined with the relative concentration of the Corps program in the southern states, the study in all probability overstates the net impact figures to some, it is judged, rather small extent.

tion of the technique proposed by H. M. Groves is accepted.<sup>13</sup> Through this approach, the proportion of the total federal tax burden which is borne by each state for any given year can be estimated. By assuming that a state's residents contribute to any federal program in proportion to their contribution to the total federal tax take and that the marginal and average federal tax burdens to any given state are equal, the necessary charges paid by a state's residents in support of Corps construction in any given year can be estimated.

The concept of the individual state's share of future federal operation costs arises as the federal government, in constructing certain projects, commits itself to an annual expenditure of operating charges for the life of these projects. In any given year, therefore, the federal commitment is increased by the present value of the future operating charges for those projects on which construction was commenced in that year.<sup>14</sup> Thus, for any given state, the burden of this annual commitment is the share of the federal present value which it must bear. The modification of Grove's technique was again used to estimate this commitment.

The following, slightly simplified, example will illustrate specifically

<sup>13</sup> See Harold M. Groves, *Federal, State and Local Government Fiscal Relations*, Senate Document 69, 78th Cong., 1st Sess. (Washington, D.C.: GPO, 1943), pp. 207-20. In estimating the state incidence of federal taxes by this method, each individual tax is placed into one of three classifications, depending on the assumptions made in its allocation. First, the individual income tax and the gift, estate and admission taxes are allocated to the states on the basis of collection (assuming that the taxpayer files the return in the state in which he resides). Second, the corporate taxes—income, capital stock and excess profits—are distributed in proportion to income from dividends and interest. This procedure is based on the orthodox assumption that the corporate taxes affect the corporation directly and thus reduce dividend payments by the amount of tax collected. This assumption has been challenged by the hypothesis that some part of the tax may be shifted forward to the consumer. At present, the problem appears to be statistically insoluble. See J. Fred Weston, "The Incidence and Effects of the Corporate Income Tax," *Nat. Tax J.*, Dec., 1949, pp. 300-15. Finally, all of the remaining taxes are distributed on the assumption that the final consumer bears the burden of the tax: motor vehicle taxes in proportion to state motor vehicle registrations, alcoholic beverage taxes in proportion to the state retail liquor sales and the remaining taxes, including retail excise taxes, in proportion to state retail sales.

<sup>14</sup> In estimating the present value of the annual federal commitment the following approach was used. First, the arithmetic mean of the ratio of federal operating costs to estimated federal investment cost weighted by federal construction expenditure was computed from Corps data for the entire body of projects constructed in the South since the war; that is,

$$\bar{X}_w = \frac{\sum \left[ C_f \cdot \left( \frac{O_f}{K_f} \right) \right]}{\sum C_f}$$

when  $C_f$  is the federal construction expenditure for a given project and  $\left( \frac{O_f}{K_f} \right)$  is the ratio

of federally-borne annual operating costs to estimated federal investment costs. This mean ratio was then accepted as the best guess of the ratio for the entire Corps program during this period. From this mean ratio (which was .008 in the study), the present value of the annual federal commitment of future operating costs was determined for each year's federal construction appropriation. This present value figure was, finally, taken as an estimate of the expected cost to the body of taxpayers at time  $t$  (the time of the construction appropriation) of this additional commitment.

the approach used in the study. The question which is posed is the following: In any given year (say 1959) how did a given state (say North Carolina) fare when the value at time  $t$  (1959) of what it received from the program is compared to the value at time  $t$  (1959) of what it contributed? In 1959, North Carolina received construction appropriations on only one project, the Manteo Bay project. The total estimated federal cost of this project was \$1,145,000 but \$1,339,000, or 117 percent of the estimated cost, was actually spent in constructing the project. The benefits to be yielded from this project were estimated by the Corps to be \$228,485 annually for a period of fifty years, which when discounted at an annual rate of 4.5 percent yields a present value of \$4,515,000. In 1959, \$722,000 was appropriated or 54 percent of total appropriations. From this data then, the present value of benefits in 1959 was estimated to be \$2,645,000. This value represents the estimate for the  $P(R_z)_{1959}$  term in the redistribution function. For the  $P(C_z)_{1959}$  term, the following items make up the total estimate: 54 percent of total local construction costs or \$6,000, 54 percent of the present value of local operation, maintenance, and repair costs or \$4,000, 1.256 percent (North Carolina's share of the total federal tax burden) of total Corps General Construction appropriations in 1959 or \$7,804,000 and 1.256 percent of the present value of the additional federal commitment for future operating costs in 1959 or \$1,234,000—all of which totals \$9,048,000. Thus, it is concluded that, in 1959, North Carolina experienced a net benefit of —\$6,403,000 from the program, the amount by which  $P(C_z)_t$  exceeds  $P(R_z)_t$ .

From both the theoretical and illustrative discussion of the net benefit function then, it is clear that a numerous and complex set of forces interact to determine the regional impact of a federal expenditure program. In attempting to summarize some of the more obvious of these forces, let us briefly denote the major determinants of the two elemental terms of the net benefit function. First, consider the gross benefit term of the equation. For any region, this term depends upon two primary variables: (1) the amount of the appropriation received by projects in that region during a period of time and (2) the productivity of those projects. That is, the gross value of the program to a region is a function of both the number and size of the projects constructed in the region and the returns per unit of cost of those projects. Second, the major determinants of the cost term of the equation must react through the four subterms of which total regional cost is composed. These subterms are: (1) the initial charges borne locally, (2) the present value of future charges borne locally, (3) the region's share of federal taxes in support of national Corps construction appropriations, and (4) the region's share of the present value of the federal commitment of future opera-

tion costs of all projects constructed under the Corps program. Thus, the major abstract determinants of the size of this composite term are some combination of the size of the total Corps program, the income level of the region in question (upon which the regional federal tax burden depends), the proportion of both initial and future costs which the region bears for its own projects, and, finally, the proportion of future costs which the federal government bears on all of the projects which it constructs under the program.

*Analysis of Regional Impact.* The results secured from this approach applied to all of the 165 projects are presented in Tables 1 through 3. Using these results, three facets of the program's regional impact will be discussed. First, the absolute gross and net benefits and costs of the program to the ten southern states will be presented; second, the gross and net per capita benefits and costs to these states will be analyzed; and, finally, the trend of net benefits over time to the entire region will be discussed.

From the evidence in Table 1, it is clear that a substantial net redistribution of income in favor of the South has taken place through the Corps program. As the figures demonstrate, the General Construction appropriation of about \$1.5 billion (see Table 3) to the ten states yielded about \$2.3 billion in gross benefits or about 150 percent of the value of the appropriation. After deducting the necessary program charges of nearly a billion dollars which were borne by the states, the net impact of the program on the southern region totals something over \$1.3 billion.

As for individual southern states, the results are extremely varied. In both gross and net benefit estimates, Kentucky appears to have been the state most favorably affected, displaying a gross gain of \$500 million

TABLE 1  
GROSS AND NET RETURNS AND COSTS OF THE CORPS OF ENGINEERS GENERAL  
CONSTRUCTION PROGRAM TO TEN SOUTHERN STATES, 1946-62

(Thousands of Current Dollars)

State	Gross Returns	Costs	Net Returns
Alabama.....	156,253 (7)	76,459	79,794 (7)
Arkansas.....	337,167 (3)	47,319	289,848 (2)
Florida.....	341,312 (2)	218,190	123,122 (6)
Georgia.....	232,799 (5)	107,118	125,681 (5)
Kentucky.....	494,651 (1)	98,012	396,639 (1)
Louisiana.....	307,134 (4)	112,692	194,442 (3)
Mississippi.....	15,160 (10)	39,039	-23,879 (9)
North Carolina.....	60,610 (9)	116,394	-55,784 (10)
South Carolina.....	127,070 (8)	48,145	78,925 (8)
Tennessee.....	229,794 (6)	98,584	131,210 (4)
South.....	2,301,950	961,952	1,339,998

and a net gain of \$400 million. The residents of two lowest states, Mississippi and North Carolina, actually contributed a larger amount to the program than they received from it, experiencing losses of about \$24 million and \$56 million, respectively. The three states which reaped the highest net benefit (Kentucky, Arkansas, and Louisiana) experienced a total net gain of nearly \$900 million or two-thirds of the total for the ten-state area. On the other hand, the lowest three states (South and North Carolina and Mississippi) taken together experienced a net loss with only South Carolina receiving a positive impact. Finally, the case of Florida presents an interesting example of the divergence between net and gross impact. Although it ranked second in terms of gross benefits with an estimate of over \$340 million, it ranks sixth in terms of net benefits with nearly two-thirds of the gross benefit figure absorbed in necessary charges. Because a large proportion of the total state costs are absorbed by federal tax commitment (which is itself based on state income), Florida ranks relatively low in relation to the rest of the South in net impact, it bearing nearly 25 percent of the total southern tax burden.

Turning from the distribution of absolute benefits to the distribution of per capita effects as presented in Table 2, a similar yet somewhat different picture is seen. Again, it is evident that the southern region

TABLE 2  
PER CAPITA GROSS AND NET RETURNS AND COSTS OF THE CORPS OF ENGINEERS  
GENERAL CONSTRUCTION PROGRAM TO TEN SOUTHERN STATES, 1946-62  
(Current Dollars)

State	Per Capita Gross Returns	Per Capita Costs	Per Capita Net Returns
Alabama.....	49.47 (8)	24.27	25.20 (8)
Arkansas.....	182.55 (1)	25.61	156.94 (1)
Florida.....	87.27 (4)	55.79	31.48 (7)
Georgia.....	62.92 (6)	28.95	33.97 (6)
Kentucky.....	164.77 (2)	32.65	132.12 (2)
Louisiana.....	102.86 (3)	37.74	65.12 (3)
Mississippi.....	6.97 (10)	17.95	-10.98 (9)
North Carolina.....	14.06 (9)	27.00	-12.94 (10)
South Carolina.....	56.33 (7)	21.35	34.98 (5)
Tennessee.....	66.82 (5)	28.67	38.15 (4)
South.....	74.77	31.26	43.51

experienced a substantial positive impact from the program. Ignoring the costs of the program, the South received nearly \$75 per person. After accounting for the costs of participating in the program, the net per capita effect is greater than \$43.

In analyzing the per capita impact on individual states, Arkansas is seen to be the top state with gross returns of \$182 per person and net



gains of \$157 per person. On the other hand, the effect of this federal program actually cost the average resident of North Carolina \$13 and the residents of Mississippi an average of \$11 per person. Also, it is interesting to note that, on a per capita basis, only the top three states have benefited more than the region as a whole while the seven lowest states have a net benefit figure below the entire region. Finally, the effect of the relatively greater federal tax burden carried by Floridians is again evident in Table 2, as per capita costs of the program to Florida residents are double the per capita costs to the rest of the region. Thus, whereas Florida ranks fourth in the gross per capita benefit figure, it ranks seventh when allowance is made for necessary charges.

The postwar trend in the level of appropriations and the gross and net impact of the program on the South is shown in Table 3. The effects of the Eisenhower policy of "no new starts" and the curtailment of expenditures during the Korean war are clearly visible. From 1950 to 1959, none of the series display any trend whatsoever. Only in recent years has the program regained the rapid growth which it displayed in the immediate postwar period.

Thus, the use of this income redistribution function has defined three important phenomena. First, it is clear that the South has experienced a substantial and favorable redistribution of income as evidenced by the positive net benefit figures. Second, this impact has affected individual states in an extremely diverse manner, with some states reaping a large

TABLE 3

APPROPRIATIONS AND GROSS AND NET RETURNS TO THE SOUTH FROM THE CORPS OF ENGINEERS GENERAL CONSTRUCTION PROGRAM, BY YEAR, 1946-62

(Millions of Current Dollars)

Fiscal Year	Appropriations	Gross Returns	Net Returns
1946.....	25.6	32.6	16.8
1947.....	47.0	66.5	44.0
1948.....	55.6	82.9	43.8
1949.....	86.8	126.0	73.3
1950.....	100.0	150.9	92.3
1951.....	102.6	154.6	93.2
1952.....	91.9	146.5	88.0
1953.....	69.3	99.1	47.9
1954.....	53.2	71.2	36.8
1955.....	58.1	88.2	49.6
1956.....	70.5	11.2	56.1
1957.....	80.2	131.6	73.5
1958.....	68.6	106.0	47.4
1959.....	112.1	163.2	85.8
1960.....	150.7	229.9	142.9
1961.....	165.0	268.9	170.3
1962.....	169.7	272.6	175.3
1946-62.....	1,506.9	2,301.9	1,340.0

positive impact and some states experiencing losses. Finally, it has been shown that the trend of both appropriations and net benefits to the South has been upward with the decade of the 1950's being one of remarkable stability.

In conclusion then, it is hoped that the technique treated in this paper may help both to identify and, more important, to evaluate those elusive variables which determine the regional impact of federal expenditures. By "reverting" to rather straightforward direct estimates of both regional costs and returns, the present method strives to avoid those problems present in the approaches cited above; namely, the problem of wide error margins due to the use of inappropriate aggregate measuring devices or inadequate regional input-output coefficients and the problem of excessive subjectivity which is introduced by rather informal assessment procedures. Although subject to drawbacks of its own, it is suggested that the present approach may provide a more satisfactory technique than those previously utilized.

# FARMER RESPONSE TO PRICE IN A SUBSISTENCE ECONOMY: THE CASE OF WEST PAKISTAN\*

By WALTER P. FALCON  
*Harvard University*

## *Introduction*

The extent to which farmers in underdeveloped areas respond to price changes is a widely debated subject. To date, however, most discussions have been speculative in nature, and have been based on a priori consideration of peasant behavior and institutional limitations.

Though a few authors [4] [14] have argued that farmers in poor countries respond positively to price and income incentives, the majority [12] [13] [16] [18] appear to believe in negative or zero price elasticities of supply. The idea that farmers do not respond to economic incentives appears repeatedly in the structuralists' writings on inflation; it is also a central theme for those who favor economic development via industrialization and high agricultural taxation. Finally, a zero or negative response often is claimed by those who favor greatly expanded shipments under P.L. 480 auspices.

All of the above are key policy issues, and the purpose of this paper is to examine more thoroughly and more quantitatively the direction and magnitude of farmer responses to price in a low-income economy. The discussion here is in the context of West Pakistan, though many of the findings appear to be applicable to a much wider area. The major thesis set forth is that farmers in West Pakistan, when given the opportunity, do respond to price and income incentives.

## *Disagreements on Price Response*

Before turning to the empirical results for Pakistan, it is important to analyze the source and nature of the disagreements on price response. There is no simple explanation that can be given to account for the varying opinions, though the confusion seems to stem mainly from three general sources.

The first set of reasons arises from conceptual difficulties in defining price and quantity variables. In large part, this confusion has been created by analysts and policy-makers who have not distinguished

\* This paper is a preliminary report of a much larger and more thoroughly documented forthcoming study entitled, "Determinants of Agricultural Production in West Pakistan." The analysis draws heavily on my Ph.D. thesis [5], on fieldwork done in Pakistan between 1961 and 1963, and on continuing work being done under the auspices of the Harvard Center for International Affairs. I have benefited greatly from the helpful comments of Edward Mason, Robert Dorfman, and Gustav Papanek, though they bear no responsibility for any deficiencies that may remain.

clearly (1) between the elasticity of aggregate output (all commodities) and the elasticity of a single commodity, (2) between the farmers' allocation of land and nonland resources in response to price, i.e., between "acreage" and "yield" responses, (3) between elasticities for "cash" and "home-consumed" crops, (4) between "production" and "marketing" elasticities, and (5) between the varying lengths of time to which a supply elasticity may refer. Most of these distinctions are obvious and are not new to the literature on supply analysis. Nevertheless, a review of the literature on underdeveloped countries reveals that these distinctions are not being made. There is, instead, a tendency to talk about a single supply elasticity for agriculture. Since even for a given agricultural commodity there may be several types of responses relevant to policy-makers, it is both confusing and meaningless to refer to one supply elasticity for agriculture.

Statistical problems in the estimation of supply elasticities are another source of many of the wrong conclusions on farmer response. Part of the difficulties are the result of limited and unreliable data, though this cannot be the entire explanation. Some of the controversy can be traced also to the form of the supply model used for estimation purposes. For example, elasticities obtained from simple supply models, i.e., quantity or acreage as a function of price lagged one year, are smaller usually than those derived from distributed lag formulations.<sup>1</sup> However, even after allowance has been made for this range of choice in models, there have been clear cases where the analyses were wrong methodologically. Neglect of the identification problem, the inclusion of "abnormal" observations, a disregard for problems of aggregation, and a failure to deflate prices properly account for many of the varying results and wrong conclusions.<sup>2</sup>

Micro surveys of individual cultivators also have been at fault in producing inconsistent answers [e.g., 3]. Many questions on price response have been worded in such a way so as to assure a particular answer; e.g., to ask a farmer if he would grow more wheat if the price of wheat were raised is to invite a "yes" answer. Also, these surveys have not solved the problems of inquiring about responses to price under conditions of price and yield uncertainty. Nor have they always been successful in establishing rapport with villagers who (often for good reason) view outsiders with considerable suspicion.

<sup>1</sup> See Nerlove [17] for details of the distributed-lag method. For example, Krishna [15] found an insignificant price elasticity for cotton in the undivided Punjab (1924-43) with a simple lagged model. However, with a distributed-lag formulation, he found short- and long-run elasticities of .34 and .75, respectively.

<sup>2</sup> See, for example, Board of Economic Inquiry (Punjab) [11]: "The positive correlation [between undeflated prices and nonstandardized acreage] in the case of rice, wheat, bajra, maize, gram, sugarcane and cotton (American) indicates that higher prices result in an expansion of area and vice versa, while the negative correlations of barley, jowar, repeseed, cotton (desi) and tobacco mean that price and acreage run in opposite directions" (p. 43).

A third set of misunderstandings on supply response stems from the group who believe that "illiterate" and "irrational" are synonymous words when applied to peasant cultivators. This need not be the case, as we will see in the next section.

### *Empirical Results for West Pakistan*

It is not easy to measure relevant supply elasticities for West Pakistan. The time series data on yields, acreages, marketings, and prices, that are a virtual necessity in elasticity computations, are partial in coverage and only moderately reliable [6]. There is inadequate information also on storage, transportation, credit, and other institutional constraints. Nevertheless, it is possible to set forth several farmer-reaction hypotheses and to test some of them against the limited aggregate data.

*Some Hypotheses.* In West Pakistan farms are very small and fragmented. Government services, especially credit and extension facilities, are meager. Government officials, outside conquerors (of whom there have been many), and the "market economy" all are treated with suspicion. Rural diets are inadequate in quantity and composition, and for the bulk of cultivators there is a continual threat that farm production will fall below the consumption needs of the farm family. As a result of these and other factors, uncertainty minimization and food grain self-sufficiency have been the historic keys to survival, and are the major factors in farm planning. The force of these two elements results in a relatively low-valued, subsistence-oriented cropping pattern for the majority of small farmers.<sup>3</sup>

It is not sufficient, however, to explain this phenomenon as simply the result of an ignorant, tradition-bound small peasant cultivator. Though traditions may limit the latitude of choice open to a small cultivator, he is probably no more ignorant than the large farmer when it comes to formulating measures of expected yields and prices; it is just that his ignorance costs him more in a real sense. For if a cultivator produces food grains for his own needs, he runs the risk of a poor crop. If he produces a cash crop and leaves part of his food needs uncovered, he still runs the risk of a poor crop as far as the cash crop is concerned. He also runs a double risk on prices. The price of the cash crop may have fallen by the time he is ready to sell it, and the price of the food grain may have risen by the time he is ready to buy it. Adverse price movements—and 100 percent seasonal and annual movements are not uncommon—are very real costs to the cultivator.

<sup>3</sup>For example, I have just completed a linear programming analysis for the "Revelle Report" [20] which shows that a 50 percent increase in the value of net output could be obtained in southern West Pakistan by changing the agricultural production pattern (holding prices, inputs and technology constant).

Hence for any small farmer, the risk increases at an increasing rate the higher the proportion of the household consumption requirements that must be purchased. Given the asymmetry in effect between gains and losses in income, farmers continue to grow food grains and low-risk crops.

If the above is a reasonably accurate picture of the planning model used by the majority of farmers in West Pakistan, several types of short-run macro price responses can be postulated:

First, there is likely to be little correlation between the total acres planted to all crops and the terms of trade for agriculture.<sup>4</sup> Population pressure, the low level of rural consumption, the small portion of crops grown for the market, a lack of irrigation water, and the small size of most farms suggest no reason why there should be any logical association in an agriculture where very poor land is already being farmed. Climatic and reclamation factors are likely to far outweigh fluctuations or increases in acreage caused by the changing terms of trade for agriculture.

Second, there is likely to be a considerable range over which there is no cash crop-food grain acreage substitution in response to changes in the relative price between food and cash crops. Because of the uncertainties of the market and a historic (and rational) desire on the part of most farmers for having wheat in their *godowns* (storage bins), the consumption requirements are likely to be grown on the farm. To the extent that there is a price response, it is likely to occur in years when the food grain stock position is good and/or to be restricted to those cash crops whose returns are sufficiently high to overcome the risks of not meeting home consumption needs.

Third, there is likely to be a positive acreage response to relative prices between crops grown for the market. These would be produced on the "residual" acres of the small farms, i.e., total acres minus food acres, and on the larger farms. However, uncertainty again may play a large role when a new cash crop is introduced.

Fourth, there is likely to be little year-to-year change in relative yields in response to expected relative price changes between food and cash crops, or even between two cash crops. In addition, there is likely to be little relationship between changes in yields and changes in the prices of agricultural outputs relative to the price of inputs. Severe capital rationing, the fixed and limited nature of most resources, physical unavailability of many factors, and ignorance of better farm-

<sup>4</sup>The "terms of trade for agriculture" is defined in this paper to mean the price of agricultural outputs relative to the price of agricultural inputs. (Note that the denominator does not include consumption goods.) "Relative price" is defined as the ratio of prices between different agricultural outputs; e.g., the price of cotton divided by the price of wheat.

ing methods militate against such responses. Major yield changes are caused more significantly by climatic variations, waterlogging and salinity, pests, etc.

*Quantification of the Hypotheses.* Because of a lack of data, especially information on "prices paid by farmers," all of the preceding hypotheses cannot be tested. It is possible, however, to make some quantitative assessments of the cotton and wheat "acreage" elasticities, and of the cotton "yield" response. On the basis of the wheat and cotton analyses, it is possible also to make some judgments about a supply elasticity for aggregate agricultural production. Because of limited space, only relatively short-run (year-to-year) changes are analyzed.

*An Acreage Model for Cotton.* For a cash crop such as cotton, it can be hypothesized that acreage in any given year is related to the Pakistani farmers' expectations of the relative profitability of the crop.

Further specification of the model is not an easy task—partly because of the variety of crops that can be grown in West Pakistan. It is costly in terms of degrees of freedom to include the prices of all the crops that are potential production substitutes for cotton; however, more serious than this loss is the problem of multicollinearity among the price variables. Because of these complications, a single price series was formed: the harvest price of cotton divided by a weighted average of the production substitutes' prices. The ratioing automatically deflates the price data.

A further problem also exists in defining "expected" price. One assumption that can be made about farmers' reactions is that the relative price of the previous period is used as the best estimate of the forthcoming sale price. This assumption, or any other which suggests that expected prices are related to the actual prices of preceding prices, may seem unduly restrictive; however, it is necessary in any single equation approach to the measurement of supply elasticities.

Hence, in the model actually fitted, the percentage change in cotton acreage was regressed against the lagged relative price of cotton. This equation was estimated for the eight major cotton producing districts of former Punjab for the period 1933-34 to 1958-59.<sup>5</sup>

$$(1) \quad \left( \begin{array}{c} \% \text{ change in 8-district cotton acreage} \\ t - 1 \text{ to } t \end{array} \right)$$

<sup>5</sup>The districts include Lahore, Shekhpura, Gujrat, Shahpur, Montgomery, Lyallpur, Jhang, and Multan. The acreage observation for 1943-44 was omitted because of the price distortions of 1942-43 caused by World War II. For further details, see Falcon [5], Chap. III.

$$\begin{array}{c}
 = -41.1 + 16.7 \\
 (2.4)
 \end{array}
 \left[ \begin{array}{c}
 \text{Price of cotton} \\
 t - 1 \\
 \hline
 \text{Weighted price of rice, bajra, jowar,} \\
 \text{corn and sugarcane} \\
 t - 1
 \end{array} \right]$$

$$R^2 = .70$$

$$n = 25$$

The estimates given in equation (1), show that 70 percent of the variation in cotton acreage is explained by this two-dimensional model. For an agricultural supply analysis the estimating equation "fits" the data remarkably well. This is particularly true since the dependent variable has been differenced. Using the slope coefficient of (1), the mean supply elasticity for cotton in terms of acres is .41.<sup>6</sup> A short-run cotton elasticity as high as .4 appears quite remarkable in light of similar studies for other countries. Brennan [2] for example, found coefficients of only about .35 for cotton in the United States.

Another set of measurements which is useful for testing several of the hypotheses involves particular cross-elasticities of supply. Because of the problem of multicollinearity, none of the multiple-regression formulations that used individual crop prices proved entirely successful. However, these models did show a significant negative relationship between cotton acreage and the price of sugarcane (the other major cash crop). The food grain price coefficients were all negative in sign but were not statistically significant. Both of these results appear consistent with the food grain self-sufficiency idea expressed earlier in hypotheses two and three.

Two other aspects of the cotton acreage model should be noted. First, a number of surveys have shown that the ratio of cash crop acres to total farm acres increases with the size of the farm [e.g., 7]. This further suggests that increased population pressure and reduced farm size, along with the problems of waterlogging and salinity, are major reasons for the stagnation and decline of Punjab cotton production in the last twenty years. However, these are longer-run phenomena which cannot be analyzed fully in this paper.

Second, the food grain stock position may have much to do with the acreage planted to cotton. For example, the harvest of fall-planted

<sup>6</sup> Justification for the form of equation (1) is given in Falcon [5], Chap. III. The elasticity was computed by multiplying the slope coefficient (16.7) by the mean of the price series (2.44) and dividing this product by 100.



(*rabi*) cereals was very low in the Punjab during 1952-53. In the following spring season more of the acres went into summer food grain production and less into cotton than otherwise would have been expected.<sup>7</sup> This is a further indication of the importance that farmers place on food grain self-sufficiency.

*A Yield Model for Cotton.* Just as a farmer might react to relative prices by growing more acres of a given crop, he might react also by allocating his nonland resources so as to increase his yield per acre. Though this phenomenon has been of considerable importance in the United States, it has been much less important in West Pakistan.

When percentage change in cotton yield was substituted as the dependent variable in the formulations of the acreage section, not a single estimating equation yielded an  $R^2$  of greater than .1. For example, the ratio model that explained 70 percent of the year-to-year changes in acreage gave an  $R^2$  of only .002 when change in yield was used as the dependent variable.

On the basis of these results, there is good reason to reject, at least for the time being, the hypothesis that expected relative prices have had a significant effect on changing cotton yields in West Pakistan. Even when other variables (such as water availability, salinity, etc.) were "partialled out," yields and lagged prices were not associated. Nor was there a correlation when various corrections were made on the yield data that are thought to be biased [6].

There appear to be a number of reasons for this lack of correlation and the seemingly inconsistent results between the acreage and yield results. Impressions gained on field trips into the Punjab suggested that physical unavailability of inputs, capital rationing, limitations imposed by uncertainty, ignorance of better methods, and a lack of incentives were among the most important.

The above observations are supported in part by a recent National Sample Survey of West Pakistan cultivators [8]. For example, 54 percent of the cultivators sampled did not know of the Agricultural Department's program of improved seed; 70 percent had never visited a demonstration plot; 68 percent were not aware of the plant protection services of the government; 96 percent of the sample cultivators borrowed during the year with over 70 percent of funds coming from friends and relatives; only 4.3 percent of the cultivators used the services of the extension staff.

With regard to fertilizer, the per acre use over the last twenty years has been limited sufficiently to have had, at most, only a marginal

<sup>7</sup> If food grain stocks and food grain prices are negatively correlated, some of the "stock" effect will be reflected in the price model.

effect on aggregate average yields [9]. Physical availability rather than price has been the major consideration in its use, and when available, fertilizer and insecticides have been used by those farmers with a good cash position. However, in those instances where fertilizer was applied, reports [e.g., 11] show that the fertilizer was used usually on the cash crop with the highest expected net return.

Water—a very constraining factor in West Pakistan—is even less subject to control by the farmer. The quantity of rainfall that falls is completely outside the realm of farm planning. The amount of irrigation water applied is likewise more a function of its availability than of the cost of the water because its marginal value productivity is much higher than its cost to the farmer [10] [20].

Labor is another factor that might be shifted in response to expected prices. No studies have been conducted on this important topic and major reliance must be placed therefore on indirect evidence. However, one is impressed by the fact that the number of ploughings, hoeings, etc., seem to be fixed rather rigidly for any given region. Even when shifts are made (say from broadcasting to line sowing, which requires slightly more labor), they tend to be irreversible. Thus labor responses might have accounted for a trend in cotton yield (if there had been a trend), but not for year-to-year variations in it.

Institutional constraints also may affect the allocation of resources and the adoption of new techniques. Tenure arrangements are one such example. If none of the increases in productivity accrue to the tenant, there is no incentive for the peasant to innovate or to work harder.

Hence, the lack of “yield” response on the part of West Pakistan cultivators is explained, not on the grounds of economic irrationality, but rather by a lack of incentives, by capital rationing, by limited inputs, and by ignorance of better methods. Given this explanation, the conclusions on cotton acreage and yield responses are entirely consistent with one another.

*An acreage Model for Wheat.* A discussion of wheat provides an interesting analytic complement to the previous discussion of cotton. Wheat is the major food grain in West Pakistan, and there is a high component of home consumption—perhaps as much as 75 percent. In addition, wheat is grown under more varied geographic conditions.

There are two aspects of the wheat model that should be mentioned at the outset. Because of a lack of data on wheat marketing, all of the discussion must be with regard to production. This is an important distinction, because prices may affect the amount marketed, even if they do not affect the quantity produced (or vice versa). Second, there

are three distinct wheat-growing regions in the former Punjab.<sup>8</sup> In seven of the districts, wheat is grown under irrigated conditions; in four, it is a rain-fed crop; and in five districts there is the possibility of an intracrop substitution between irrigated and rain-fed wheat. Variables which explain variations in one region are not likely to do so in the other two.

Various models were tried for each of the areas. For the rain-fed region, relative prices explained very little. The total fall-planted acreage was highly correlated (.7) with rainfall variables, and regardless of relative prices between wheat and the other fall crops, 65 percent of the total area went into wheat production.

There are several reasons why the acreage fraction was always at 65 percent. Pulses and barley, which together with wheat account for over 90 percent of the total acres, are more complementary with wheat than competitive. These crops often are sown with wheat as mixed crops as a contingency against "dry" weather. Prices therefore are incidental to cropping patterns, and the acreages of these three crops move in conjunction with total acres. A more competitive relationship might have been expected between wheat and oilseeds, because the latter often are sold for cash; however, such an association was not the case. The major explanation offered is that the increase in gross return from raising oilseeds rather than wheat (at most 10-15 percent) was not sufficient to overcome the uncertainty discount of growing a cash crop.

The seven-district irrigated region provides an interesting contrast to the rain-fed region. Fewer climatic restrictions provide a wider range of cropping alternatives in the canal area. In this more developed region, there was some price response between wheat and the year-long crop of sugarcane—a crop which competes with wheat for both land and water. About one-fourth of the variation in wheat acreage can be explained by the lagged ratio of wheat to sugarcane price. The limited, but significant, short-run price elasticity of .1 to .2 is an important supplement to the previous findings. It suggests some price responsiveness even for the major food grain (1) when climatic restrictions do not severely limit the cropping pattern and (2) when the expected return from the cash crop was sufficiently high (perhaps double) to overcome the risk of not growing a food crop.

An analysis of the five-district area that contains large amounts of both irrigated and rain wheat acreages provides an important check of the effect of water availability. One of the most interesting

<sup>8</sup> The rain-fed districts are Jhelum, Rawalpindi, Attock and Mianwali. The irrigated districts are Lahore, Lyallpur, Montgomery, Jhang, Multan, D. G. Khan, and Muzaffargarh. The mixed districts are Shekhupura, Gujrat, Shahpur, Sialkot, and Gujranwala. For further details, see Falcon [5], Chap. IV.

findings was the high correlation (.7) between October-December rainfall and the ratio of rain-fed to irrigated wheat. This correlation indicates that when rainfall was greater than "average," more wheat was raised as a rain-fed crop. This gives another indication of the ability of the Punjabi farmer to react to changing circumstances.

There are several important conclusions to be drawn from the price and water analyses of the three regions. The first point to be noted is that water availability appears to be the most important variable in determining wheat acreage. In addition, water availability places varying limits on the amount of price responsiveness that is possible. Finally, the analysis shows that even for the case of food grains, there is some price response when climatic restrictions are not overly severe, and when there is a lucrative cash crop alternative.

*A Model for Aggregate Output.* Though there are insufficient data to make rigorous statements about the effect of past changes in the terms of trade for agriculture, several points seem clear. First, aggregate acreage tended mainly to vary with water availability. Given the extreme population pressure on land (and the profits to be made in industry by potential investors [19]), the terms of trade were of little significance in affecting the total acres planted. Nor did changes in the price of agricultural outputs relative to inputs appear to affect the yield of agricultural products. This is due primarily to the factor proportions and factor rigidity problems mentioned in the cotton yield analysis.

However, this is not to state that prices have no effect on aggregate agricultural output. We have seen the relative prices within agriculture affect the composition of agricultural output. Therefore, if, as a result of relative price changes, an acreage shift is made from lower to higher valued crops, the value of aggregate production will increase. There are very difficult conceptual and empirical problems to overcome in measuring the (real) values of these alternative outputs. However, considerations of this kind cannot be overlooked if relative prices within agriculture are to be deliberately changed through export taxes, price controls, etc. Furthermore, it would appear that changes in the acreage composition of output due to price changes (e.g., the increase in the high-valued crop of sugarcane) have been responsible for a small part of the growth in the total production of West Pakistan's agriculture.

### *Summary, Conclusions, and Implications*

This discussion of West Pakistan provides a number of insights on farmer response to price in a near-subsistence economy. In particular, it gives partial answers to two important points: (1) why there has

been so much confusion on agricultural price response and (2) what the potential is for breaking bottlenecks in agriculture via the price mechanism.

We have seen that much of the confusion on supply elasticities has been the result of thinking only in terms of a single supply elasticity for agriculture and of improper statistical techniques.

In addition, we have seen that even in a low-income region such as West Pakistan, there may be significant acreage responses to changed relative prices. This is particularly true for cash crops such as cotton where the acreage elasticity of  $+ .4$  equals that found for the United States. It is even true in the case of food crops when climatic conditions are not overly severe and when there are good cash crop alternatives.

While there have been significant short-run acreage shifts in response to changed relative prices, there has been little responsiveness in the allocation of nonland factors. Changes in rainfall, irrigation water availability, and other geographic factors still are the major determinants of changes in yields. However, this limited price response does not imply automatically that farmers do not respond to economic incentives. For they must be able to acquire fertilizer and better seeds, i.e., they must be given the opportunity to react, before they can be labeled as unresponsive.

These yield and acreage results have important policy implications for Pakistan. They suggest that it is possible to shift the composition of agricultural output by changing the relative prices within agriculture. They also emphasize that unless there is a thoroughgoing reform in the services and facilities made available to farmers (e.g., transportation, storage, credit, fertilizer, technical knowledge, etc.), higher prices alone can have little effect on increasing yields per acre. On the other hand, the acreage evidence suggests that farmers of the area will respond to economic incentives if given the opportunity to do so. Hence, West Pakistan's agricultural difficulties appear to be less a matter of farmer motivation and more a problem of reducing uncertainty and of improving the organization of agricultural services.

Some final words of warning must be given about the use of the quantitative results presented in this paper. The estimates were measured over a limited range of prices; they were calculated also under conditions of price uncertainty. It has been stressed that uncertainty is a major factor in farm planning, and it is likely that price responses would be higher if farmers were assured of guaranteed prices. It should be noted, also, that the elasticity estimates are aggregate measures. There are great differences in responsiveness within and between regions in West Pakistan, and these differences must be considered when policies

are being formed. Finally, these results reflect conditions of the last twenty-five years. If there is increased commercialization and agricultural development, the estimates may be inappropriate for the future.

## REFERENCES

1. Board of Economic Inquiry (Punjab), *Agricultural Statistics of the Punjab, Pakistan, 1901-2 to 1946-7* (Lahore, 1950).
2. M. J. Brennan, "Changes in Cotton Acreage in the Southeast—Implications for Supply Functions," *J. of Farm Econ.*, Nov., 1958, pp. 835-44.
3. Dorothy Cook (advisor), "Survey of Wheat Production Problems in the Village-Aid Areas of the Lahore Region," AID monograph, Karachi, Nov., 1958.
4. Rex Daly, "Appraisal of Pakistan's Second Five Year Plan," AID monograph, Karachi, Mar., 1960.
5. Walter P. Falcon, "Farmer Response to Price in an Underdeveloped Area: A Case Study of West Pakistan" (unpublished Ph.D. thesis, Harvard Univ., July, 1962).
6. ———, "Reliability of Punjab Agricultural Data," *Statistical Papers No. 1*, Institute of Development Economics (Karachi, 1961).
7. Government of India, Ministry of Food and Agriculture, *Studies in Economics of Farm Management in Punjab* (New Delhi, Jan., 1960).
8. Government of Pakistan, Central Statistical Office, *National Sample Survey, Second Round* (Karachi, 1960).
9. Government of Pakistan, Ministry of Food and Agriculture, *Agricultural Statistics of Pakistan* (Rawalpindi, May, 1961).
10. Government of Pakistan, Ministry of Food and Agriculture, *Survey Report on Cropping Patterns and Crop Intensities* (Rawalpindi, Aug., 1961).
11. Government of Pakistan, Ministry of Food and Agriculture, *Survey Report on Use of Fertilizer in Pakistan* (Rawalpindi, April, 1961).
12. Joseph Grunwald, "The Structuralist School on Price Stability and Development," in A. O. Hirschman, ed., *Latin America Issues* (Twentieth Century Fund, 1961).
13. D. R. Khatkhate, "Some Notes on the Real Effects of Foreign Surplus Disposal in Underdeveloped Areas," *Q.J.E.*, May, 1962, pp. 186-96.
14. Raj Krishna, "Farm Supply Response in India-Pakistan," *Econ. J.*, Sept., 1963, pp. 477-87.
15. ———, "Farm Supply Response in the Punjab (India-Pakistan): A Case Study of Cotton" (unpublished Ph.D. thesis, Univ. of Chicago, 1961).
16. Kusum Nair, *Blossoms in the Dust* (London: Duckworth and Co. Ltd., 1961).
17. Marc Nerlove, *The Dynamics of Supply* (Johns Hopkins Press, 1958).
18. R. O. Olson, "Discussion: Impact and Implications of Foreign Surplus Disposal on Underdeveloped Economies," *J. of Farm Econ.*, Dec., 1960, pp. 1042-45.
19. Gustav Papanek, "The Development of Entrepreneurship," *A.E.R.*, May, 1962, pp. 846-58.
20. White House—Interior Panel, *Waterlogging and Salinity in West Pakistan* (Washington, D.C., forthcoming).

## DISCUSSION

CHARLES C. HOLT: Mr. Sparrow's paper on the determinants of volume in a probabilistic market is the second of two papers, the first of which deals with expected price and the second with expected quantity flow. This bisection is highly artificial and ideally should be remedied before their publication.

In recent years economists have struggled to develop and redirect economic theory in order to increase its relevance to a dynamic and uncertain world. Mr. Sparrow has formulated a probability model of basic supply, demand, and price adjustment relationships as a means of improving our insight into the role of uncertainty in the operation of competitive markets. By the use of queuing theory and numerical methods, he has succeeded in obtaining the stochastic equilibria of a number of quantitative models. This is no mean feat in itself, considering the complexity of such stochastic processes even for relatively simple systems. But beyond this, he demonstrates an admirable ability to cut into the heart of a complex analysis and reach usable and comprehensible generalizations. As an approach to the formulation and solution of theoretical problems in this area, Mr. Sparrow's paper merits high praise.

A few technical points might be made. Item 6 under "The Intuitive Model" assumes negligible probability of neither excess supply nor excess demand. This is certainly not true for his discrete model and a footnote at least is called for. The addition of a fairly obvious term to equation (1) should take account of this case without significantly affecting his results.

In analyzing the effect of the strength of the price-change mechanism on expected volume, he sees no clear conclusion from his analysis. Yet experimentally he finds unequivocal evidence that strengthening the price-change mechanism increases the expected volume. However, a close examination of his analysis indicates that increases in price-change strength increase the volume when that strength is weak, and similarly increases in strength bring increases in volume when the strength is strong. Thus the experimental result is clearly anticipated by his own analysis although admittedly exceptional cases may occur.

When one considers what markets might satisfy the premises of this analysis, it is difficult to point to any real world markets with confidence. The stock market is mentioned in the first paper and in that market some dynamic fluctuations undoubtedly can be traced to queuing phenomena. The picture of a queue of goods (or a queue of customers) waiting to carry out a transaction with the first customer (or good) to arrive without any price restriction on the transaction is a bit strained. Orders to buy or sell stocks "at market" is the only example that occurs to the discussant. The price which presumably governs all transactions and regulates the flow of customers and goods into the market is changed in response only to the excess supply queue or the excess demand queue.

The mathematical convenience of the Markov process undoubtedly explains this formulation and this consideration can be accepted as valid up to a point.

Unfortunately, the economic conclusions may be significantly affected by the particular assumptions of the model. For example, price might be made dependent on the rate of increase of the queue as well as its size. This might significantly change some of the conclusions.

Ideally we should do a good deal of research on detail market processes so that models can fairly accurately reflect the particular stochastic processes that are involved. We can anticipate that the mathematical solutions are likely to be unavailable; so we may have to resort to queuing simulations as is common in machine scheduling studies. Indeed it is interesting to note that Mr. Sparrow finds his exact mathematical solution for the determinants of volume to be sufficiently opaque that he presents a simplified approximate analysis which offers more insights.

The assumption of the Poisson distribution has some empirical support in the sales forecasting literature. For example, see Chapter 15 of *Planning Production, Inventories and Work Force*, by C. C. Holt, F. Modigliani, J. F. Muth, and H. A. Simon (Prentice-Hall, 1960).

One would like to see studies similar to Mr. Sparrow's carried out on related problems. We need to know more about how market processes approach their stochastic equilibria. If undesirable fluctuations of markets for equities and primary products are to be reduced, we need to know more about their stochastic processes. Perhaps needed light can be thrown on Phillip's wage adjustment relation by studies of the stochastic processes of labor markets.

Mr. Sparrow has made an important contribution in showing the relevance of queuing concepts to market phenomena, and particularly in pointing up the identification problem that can occur when supply and demand interact stochastically. A great deal more needs to be done along the lines that Mr. Sparrow has pursued but increasing stress should be laid on the empirical base underlying such theoretical analyses.

D. GALE JOHNSON: We are indebted to Mr. Falcon for adding to our knowledge of the economic behavior of farmers in the poor nations. The presumption is all too general that poor and illiterate farmers are influenced by tradition and noneconomic factors and are not responsive to price and income incentives.

With a very simple estimating equation, Falcon found that 70 percent of the variance in the annual percentage change in cotton acreage in the former Punjab was explained by the lagged relative price of cotton. The supply elasticity at the mean was 0.41, a remarkably high elasticity for a response apparently based on a simple one-year lag of relative prices.

Falcon's results might be considered to be a statistical fluke were it not for the fact that there are at least two other studies that give very similar results, even though somewhat different estimating techniques were used and the time periods differ. I refer to a study by Raj Krishna<sup>1</sup> which is referred to by Falcon, and a study of jute production by L. S. Venkataramanan.<sup>2</sup>

<sup>1</sup> Raj Krishna, "Farm Supply Response in the Punjab (India-Pakistan): A Case Study of Cotton" (Ph.D. dissertation, Dept. of Econ., Univ. of Chicago, 1961).

<sup>2</sup> L. S. Venkataramanan, "A Statistical Study of Indian Jute Production and Marketing with Special Reference to Foreign Demand" (Ph.D. dissertation, Dept. of Econ., Univ. of Chicago, 1958).



Krishna's study, which covered the period from 1924 through 1943, was for the undivided Punjab and was concerned with estimating the supply response for two different types of cotton—American or so-called long-staple and short-staple cotton. The American cotton was introduced to India in the latter part of the nineteenth century but did not become important until an adapted strain became available in the second decade of the twentieth century. Based on a distributed lag model, Krishna found that the short-run elasticity of acreage with respect to price was probably between 0.28 and 0.34 and the long-run elasticity between 0.75 and 1.16. The elasticity of acreage with respect to expected yield was even greater than the elasticity with respect to price.

For the native or short-staple cotton, Krishna found a short-run elasticity of acreage with respect to expected price of 0.5 and a long-run elasticity of about 1; the elasticities of acreage with respect to yield were 0.25 and 0.5, respectively.

The relative prices used by Krishna were the prices of each type of cotton relative to a weighted average of six competing crops, of which the other type of cotton was one. The elasticity of acreage substitution between the two types of cotton was much higher than the acreage elasticities referred to above. The elasticities of the relative acreage with respect to relative prices were 0.77 for the short run and 3.85 for the long run; the elasticities of relative acreages to relative yields were much higher: 2.31 and 11.5.

Krishna undertook a similar analysis for wheat; while the price coefficient was significantly different from zero (5 percent level of significance) the short- and long-run elasticities of acreage with respect to expected price were quite low: 0.07 and 0.17.

I shall refer only briefly to the results of the jute study. In India rice is the major substitute for jute. The elasticity of acreage was determined with respect to the price of jute and the price of rice. With respect to its own price, the short-run acreage response was 0.46 and with respect to price of rice,  $-0.35$ . The long-run elasticities were higher by about a half. Venkataramanan's analysis was based upon distributed lag models.

There is one important point on which Krishna and Falcon obtained quite different answers. Falcon could find no significant relationship, either gross or net, between prices and yields for cotton. For the American cotton, Krishna found that the elasticity of yield with respect to the price of cotton relative to six alternative crops, lagged one year, was 0.65. The estimating equation from which this result was obtained included lagged irrigated area for all summer crops and weather as measured by rainfall. Thus one might argue that Falcon is being hasty when he says "there is good reason to reject, at least for the time being, the hypothesis that expected prices have a significant effect on changing cotton yields in West Pakistan." While the geographic areas are not identical, Krishna's results should not be ignored.

# EFFICIENCY IN THE TEACHING OF ECONOMICS: THE PRODUCT

## THE PROBLEM OF COMMUNICATION

By LEONARD S. SILK\*

*Business Week*

### I

What does the title of this session mean? I take it to mean what our chairman, Ben Lewis, says it means. "For generations," Professor Lewis wrote to me last April, "economists have accepted the fact that no one has ever paid any attention to us when we talk about tariffs, but I think all of us are really shocked by the quality of the public discussion on the President's tax proposals. Quite apart from conclusions, the things people say—almost everyone—bear no evidence that any economics has been taught in the present century. Well, if you will take off from here, and tell us how it looks to you and what we can do about it, we all (and I particularly) will be eternally grateful." If in fact I can say anything genuinely helpful, I shall be eternally surprised.

Professor Lewis' statement of the problem may be a little hyperbolic, but I think it is essentially correct. Whether the substance of what economists have been trying to teach is correct or incorrect, they have not succeeded in teaching it to many people.

Let me provide some evidence, taking my cue from Professor Lewis' concern over the fiscal policy debate during the past year. I think it is beyond dispute that the great majority of American economists this year have favored a tax cut to increase the nation's growth performance and to reduce unemployment. A survey of more than 500 university economists, conducted by the Chase Manhattan Bank in the fall of 1963, found that 84 percent favored an immediate tax cut, despite imbalance in the federal budget.

Yet, although the great majority of economists over the years since the Great Depression have been trying to teach the lesson that the budget need not and should not be annually balanced, the public had clearly learned little or nothing about fiscal policy. A poll of 1,588 businessmen, conducted by the Research Institute of America in the summer of 1963, found that 85 percent would favor a tax cut only provided that there be a comparable cut in federal spending.

If it were necessary to prove the point, one could cite much addi-

\*I wish to express my thanks for useful suggestions to William Wolman, M. Louise Curley, and Kenneth Boulding.

tional evidence to show the wide disparity between what the economists were teaching and what the public believed.<sup>1</sup>

Conceivably, the public was right and the economists, misled by false theories, were wrong. Even if this were true, however, it must be noted that the arguments of those opposing tax reductions, without equivalent reductions in federal spending and a balanced budget, were not distinguished for their logical consistency or economic understanding. This was even true of many leaders of the United States Congress. The well-known Rep. John W. Byrnes, of Wisconsin, Chairman of the Republican Policy Committee, attacked the Business Committee for Tax Reduction, headed by Henry Ford and Stuart Saunders, for its "retreat from fiscal responsibility" because they had not demanded equal spending cuts. To show that this was not a partisan matter, Rep. Howard W. Smith, the Democratic Chairman of the House Committee on Rules, explained his opposition to tax reduction in more homely terms: "My own background," said Rep. Smith, "is that of a small businessman who has had to scratch right hard to keep his head above water at times. The arithmetic of this thing just doesn't work out. I couldn't in my business increase my expenses and reduce my revenue and stay alive. I don't think a government can do it without an eventual change in our form of government."

I do not wish to labor the point. Yet I believe that even those professional economists who do not favor a tax cut at this time for one reason or another—perhaps because they prefer a monetary policy solution, or because they fear future inflation, or because they do not want to surrender command over future resources to the private sector, or for whatever reason—would nevertheless agree that few members of Congress or the public could handle or even follow the debate with much economic understanding.

Even if the tax bill passes in the coming session, as is probable, I think that it will be a long time before we forget the confusion and incompetence with which the tax cut debate was conducted.

I believe that the same kind of public economic ignorance manifested on fiscal and debt policy could be demonstrated in many other areas, but I shall not try to do so here.

## II

What explains the economic bewilderment and ignorance of the public? I think the first explanation is simply that economics is a dif-

<sup>1</sup> The footdragging by the House and Senate over taxes was due, at bottom, to the sensitivity of the Congress to public opinion, which was confused or apathetic over the case for tax reduction. Polls may always be criticized for their statistical flaws, but the professional politician is unlikely to take a false reading on the mood of his constituents on this kind of major issue.

difficult subject; deceptively difficult. Men who have taken doctorates in economics and worked in the field for many years often forget how difficult it is—and forget what a hard time they had in grasping its apparently simple concepts. At the same time, many businessmen and educators apparently believe that economics would not be so difficult if economists would just stop being obscurantist, pedantic, and nit-picky. Those noneconomists who would like to increase economic literacy are constantly asking economists to state some simple, basic economic truths that everyone should be taught. This is not so easy as it sounds. Let us try it:

1. There is no such thing as a free lunch.
2. The price system efficiently allocates resources, determines the bill of goods, and distributes income, in a Walrasian world.
3. When intended investment exceeds intended savings, business will rise; and when intended saving exceeds intended investment, business will decline—provided that all terms are defined in a way that will support these conclusions.
4. An increase in the money supply will cause an increase in business activity, especially if you go to the University of Chicago.
5. Real income will grow if productivity increases, provided that the real income referred to is that of the society as a whole and provided that productivity increases under circumstances in which the real income of the society as a whole is increasing.
6. Happy is the land that knows balance-of-payments equilibrium, reasonably stable but flexible prices, reasonably full employment, economic efficiency, a good income distribution in some sense, and a growth rate that is probably greater than zero, provided that this is the will of the people and provided that the will of the people is assumed to be the appropriate mechanism for determining the society's welfare function.
7. Government should do for the people only what the people cannot do or cannot do so well for themselves—all the words in this proposition being defined in any way anybody likes. (Even so, Professor Henry Wallich believes that Lincoln's classic formula overstates the case for government economic activity, contending that "it would have to be shown that the people could do something only very imperfectly, and the government very substantially better, before the government should step in."<sup>2</sup> Professor Francis Bator, on the other hand, thinks Lincoln's formula understates the case, not because government might do something less efficiently than the private economy, but because if private people want to do the wrong thing, it

<sup>2</sup> H. C. Wallich, *The Cost of Freedom* (Harper & Brothers, 1960), p. 71.

does not matter if they can do it more efficiently.<sup>8</sup> (Personally, however, I think Lincoln stated a basic truth.)

8. Competition is invariably better than oligopoly or monopoly, if the world is static, if technology is appropriate, if there are no companies with increasing returns to scale, and if you dislike socialism or realism.

9. To solve a general and nonstructural unemployment problem, either increase demand or get new unemployment statistics; to cure a noncost-push inflation, either reduce demand or get new price statistics.

10. Finally, lest you think I am cheating, I wish to conclude with the basic economic truth that value in exchange is a differential function of value in use, which simply means, as Wicksteed carefully explained, that:

... what a man will give for anything sooner than go without it is determined by a comparison of the *difference* which he conceives its possession will make to him, compared with the difference that anything he gives for it or could have had instead of it will or would make; and, further, that we are generally considering in our private budgets, and almost always in our general speculations, not the significance of a total supply of any commodity—coals, bread, or clothes, for instance—but the significance of the difference between, say, a good and a bad wheat harvest to the public, or the difference between ten and eleven loaves of bread per week to our own family, or perhaps between ten days and a fortnight spent at the seaside. In short, when we are considering whether we will contract or enlarge our expenditure upon this or that object, we are normally engaged in considering the difference to our satisfaction which differences in our several supplies will make.<sup>9</sup>

In other words, if I may try to make this a little clearer still, we consumers substitute a bit more of certain goods that now cost less than they did for certain other goods that now cost relatively more than they did, provided that our incomes are constant, that we read *Consumer Reports*, that we are not too lazy or stubborn or habit-prone or happy, that we are mercenary and do not know the people with whom we do business and have no interest in trading our high-priced dogs for their high-priced cats, that we have time and wit to make all the relevant calculations, etc. Brushing aside all such complexities, however, we may safely assert, as Wicksteed agrees, "that the differential theory of value in exchange asserts that value in exchange is value in exchange," while "all other theories assert that it is not."<sup>10</sup>

To summarize: In the great cafeteria of economic understanding—or understandings, as they say in education—there is no such thing as a free lunch.

Honest economists, including both those who have and who have not committed themselves to teaching economics to children and the gen-

<sup>8</sup> F. M. Bator, *The Question of Government Spending* (Harper & Brothers, 1960), pp. 76-98.

<sup>9</sup> P. H. Wicksteed, "The Scope and Method of Political Economy in the Light of the 'Marginal' Theory of Distribution," reprinted in R. L. Smyth, Ed., *Essays in Economic Method* (McGraw-Hill Book Co., 1963), pp. 247-48.

<sup>10</sup> *Ibid.*, p. 249.

eral public, know that economics is an unusually tricky subject; so do unusually intelligent noneconomists. As Keynes recalled:

Professor Planck, of Berlin, the famous originator of the Quantum Theory, once remarked to me that in early life he had thought of studying economics, but had found it too difficult! Professor Planck could easily master the whole corpus of mathematical economics in a few days. He did not mean that! But the amalgam of logic and intuition and the wide knowledge of facts, most of which are not precise, which is required for economic interpretation in its highest form is, quite truly, overwhelmingly difficult for those whose gift mainly consists in the power to imagine and pursue to their furthest points the implications and prior conditions of comparatively simple facts which are known with a high degree of precision.\*

The economist, on the contrary, is used to dealing with a great deal of information that comes to him in imprecise form. Even if he gets numbers that look clean, he knows that they are only shadows of a world that is anything but neat, precise, orderly, systematic. He knows that he must try to impose order on a disorderly mass of information as his normal job. Other scientists may have to do this in the beginning of their sciences or may have to do it at crucial turning points in its development, but thereafter they are filling in parts of an empirically solid structure, or in the case of mathematics, of a logically consistent structure. The economist must do essentially what an artist or writer does (though he is not so impressionistic or subjective as these): he must apprehend reality freshly every time he confronts it; he is constantly working from life, in all its buzzing, blooming confusion.

Yet the economist has a secret weapon that other people do not have. What he has that other people do not have is: economic theory. This gives him certain habits of thought that enable him to conceptualize problems that he has not seen before or problems that seem always to confront him in a new way. This is what keeps economics exciting for those who like it and what makes the economist—when he is good—such a handyman. He is able to apply his concepts to problems that, to the noneconomist, may seem totally unrelated to one another—the strategy of conflict, the farm problem, the growth of the electronics industry, the decay of a region. All such problems are challenges to the economist's ability to cope with interdependence—and poor data.

The basic bits of economic theory seem simple, obvious, even trivial. Every child realizes without being taught that if a good is cheaper, he is likely to buy more of it. But what is missing for the child is the overall system, the mode of analysis, the analogies among all types of economic activities and problems. And this is what is so hard to teach others. One gets it by doing, not by listening.

I have observed that many political, military, and business leaders

\* J. M. Keynes, *Essays and Sketches in Biography* (Meridian Books, 1956), pp. 327-28.

like to recruit and use economists, because they are good at coping with fresh problems that are not necessarily economic at all. It is hard for these admirers to understand just what's so good that economists bring to their problems. I submit that it is not merely that some economists have high I.Q.'s, but rather that they have a mode of thought that is rare among scientists or administrators or even other social scientists.

This mode of thought develops out of what I would call the economist's quasi-Talmudic training—which is long on discussion and debate, with continuous passage from the specific to the general and back again—savagely close in its textual criticism—skeptical about its own or anyone else's results—complicated and wide-ranging in its style of inquiry. My teachers Professors Calvin Bryce Hoover and Joseph Spengler impressed upon me that economics could never be a monologue art, that the economist always needed to try his reasoning on some other economist. I think this is generally true. It probably explains why the economics profession is so strong and close a fraternity; the economic monologist is not only out of touch, but always in danger of becoming a crackpot or, less seriously, a layman, if he cannot talk, talk, talk with his brethren—or, at least, read, read, read from them and to them. Preferably both.

### III

But the fraternal way that economists learn and practice their art has much to do with the failure of economists to do an effective job in educating the public.

For economists talk mostly to each other. They tend to regard the public (even ex-economists or ex-academic economists or even ex-economics department rather than business school economists) as beyond the pale. Exclusivity exists in every profession, of course, and it is, if properly exercised, a good thing. As Everett C. Hughes puts it:

Every profession considers itself the proper body to set the terms in which some aspect of society, life or nature is to be thought of, and to define the general lines, or even the details, of public policy concerning it. . . . These characteristics and collective claims of a profession are dependent upon a close solidarity, upon its members constituting in some measure a group apart with an ethos of its own. This in turn implies deep and lifelong commitment. A man who leaves a profession, once he is fully trained, licensed and initiated, is something of a renegade in the eyes of his fellows. . . . It takes a rite of passage to get him in; another to read him out. If he takes French leave, he seems to belittle the profession and his former colleagues. . . .<sup>†</sup>

While there is much to be said for these pulls toward exclusivity, since they may promote high morale and high scientific standards and values that are professional rather than worldly (this seemed to me

<sup>†</sup> E. C. Hughes, "Professions," *Daedalus*, Fall, 1963, p. 657.

•

•

the true meaning of Professor Samuelson's famous or infamous closing remarks in his Presidential address that economic scholars work "for the only coin worth having—for our own applause"<sup>8</sup>), it seems to me that, carried too far, exclusivity is particularly dangerous for the economics profession. It may cut economists off from the public they must serve. When economists venture to try to educate or influence the public, they usually find it hard to communicate; they discover that it is not easy to translate their learned arguments into comprehensible or convincing language. If I may paraphrase the great Willard Gibbs, it is worth remembering that English is a language, too.

Many economists are careless in their public utterance; they will invest endless hours of meticulous work in a journal article on some fine point of theory, and then turn around and—off the top of their heads—dictate into a machine or scribble out an argument designed to move the Congress or the general public on some vital matter of state. I have heard it said that their work in the classroom (or in preparation for the classroom) is not well or carefully done, and that the atrocious teacher who is a good researcher has no reason to doubt that he will become a full professor, but the good teacher who is not a creative researcher is wise if he leaves teaching. The really high-prestige economist may disdain to educate students at all, particularly high school students. This is perhaps understandable, both because it offers little or no professional reward and because it is painful to do—since no one has yet discovered how to do it or even whether it is capable of being done at all. Indeed, I think that some economists question whether teaching the public a smattering of economics is worth doing; many economists apparently prefer to stand on what Professor Hughes calls the common professional doctrine of *credat emptor*: let the buyer believe, let the public trust the professional to be right, since the client is not a true judge of the value of the professional's instruction or advice.

But, in economic matters, this is particularly hard for the clients to do. For one thing, the clients have strong economic convictions—and interests—of their own. For another, it is not easy for the clients to know which professional economic tutelage automatically to trust, since there is so much conflicting professional instruction and advice. Some professions—such as medicine or law—have open or tacit rules against casting aspersions upon the professional competence or recommendations of colleagues; not the economists. They frequently appear to be trying to defame one another in public. They are disputatious, within and without the cloisters, and they carry their disagreements

\* P. A. Samuelson, "Economists and the History of Ideas," *A.E.R.*, Mar., 1962, p. 18.



to the public for settlement, even exaggerating them for the sake of winning personal acclaim, while failing to get across to the public their areas of substantial agreement on important matters. Perhaps it will be said that no substantial agreement actually exists among economists on important economic topics. If that is so, it is a sorry confession about the state of economics at this juncture, and it will imply that economists have no right to try to invoke the doctrine of *credat emptor* but will have to struggle all the harder to meet the public on its own terms and to convince it of the wisdom of particular policy views. It would also imply that a greater effort should be put forth within the profession to consolidate areas of agreement, unless that would seem a premature and anti-individualist thing to do.

It will do no good to contend that the factionalism and contradictions among economists represent only a public misconception, enshrined in a hundred bad jokes. Rather, as Sir Robert Hall observed in his 1959 Presidential address to the Royal Economic Society, there are "very wide differences of opinion" between economists "on quite fundamental aspects of policy."<sup>9</sup>

Many disputes among economists, admittedly, are not about true matters of economic doctrine or analysis but, particularly on important questions, stem from differences in the political, social, or moral values of the disputants. I believe that it would help matters if economists made their values explicit rather than leaving them concealed and bending their analysis to support a value conclusion. I find myself in substantial agreement with Gunnar Myrdal's conviction that "problems in the social sciences—not only the practical ones about what ought to be done, but also the theoretical problems of ascertaining the facts and the relations among facts—cannot be rationally posited except in terms of definite, concretized and explicit value premises."<sup>10</sup> Certainly, on this matter, the public is usually not deceived. It identifies economists as far right, conservative, middle of the road, liberal, or left wing, and knows that this probably (though not necessarily) affects the nature of their analyses or recommendations.

But the economist is often self-deceived, and only succeeds in lowering his own standing and that of his profession by seeking to invoke professional authority for his personal values. I believe that when values are made explicit, we shall get better analysis, not worse, and more honest policy advice. It goes without saying that controversy is an essential part of the process by which economics is advanced and

<sup>9</sup> R. Hall, Presidential Address to the Royal Economic Society, *Econ. J.*, Dec., 1959, p. 647.

<sup>10</sup> G. Myrdal, *Challenge to Affluence* (Pantheon Books, 1963), p. vi; see also his *Value in Social Theory* (Harper & Brothers, 1958).

by which policy questions are resolved, but I believe that we have had far too much of what Profesor Devons calls "unedifying slanging matches" among economists—in which "each faction picks out those particular elements or that particular formulation of the problem which lead to the conclusion it favors."<sup>11</sup>

At the same time, economists should seek to escape from the strait-jacket imposed upon economic thought by an uncritical use of mathematical concepts and symbols. Before the big guns of the mathematicians are opened on me, let me wheel into position a big gun of my own—the mathematician, philosopher, and physicist Percy W. Bridgman. After observing in his *Nature of Thermodynamics* that there is no sharp distinction between "mathematics" and "verbalizing"—mathematics being merely a more precise language—Professor Bridgman declares that "our traditional verbal habits may have the highest guiding and constructive value."<sup>12</sup> The concepts actually used in physics (or in economics) and the operations which give them meaning are, as Bridgman says, only a few of the enormous number of conceivable concepts and operations. Bridgman says:

It is no accident that so many times we are able, by giving heed merely to our verbal demands, to evolve a concept or point of view that is relevant to an "external" physical situation. For our verbal habits have evolved from millions of years of searching for adequate methods that were not a close enough fit. The desirability of continuing to use our old verbal habits in new situations if possible is obvious enough in achieving economy of mental effort, and the probability of at least a partial success is suggested by our universal experience that absolutely sharp breaks never occur, but that a method, hitherto inadequate, can always be extrapolated beyond its present range with some partial validity. By the same token, however, the validity of any extrapolation may be expected eventually to break down, so that one may anticipate ineptnesses or inadequacies in concepts which have been formed by too uncritical a verbal extrapolation.<sup>13</sup>

Bridgman suggests that it would be a good idea to re-examine all the concepts of physics from a verbal point of view to discover how each concept evolved, whether its utility has been circumscribed by its origin, and to what extent it is misused because of its "verbal bar sinister." I feel confident that this applies with equal force to economics, probably with greater force, because, as Von Neumann and Morgenstern said in their *Theory of Games*, "our knowledge of the relevant facts of economic life is incomparably smaller than that commanded by physics at the time when the mathematization of that subject was achieved."<sup>14</sup> Yet, still foggy about many aspects of the real

<sup>11</sup> E. Devons, *Essays in Economics* (1961), p. 18, quoted in R. L. Smyth, *op. cit.*, p. 12.

<sup>12</sup> P. W. Bridgman, *The Nature of Thermodynamics* (Harper & Brothers, Harper Torchbook Ed., 1961), pp. x-xi.

<sup>13</sup> *Ibid.*, p. xi.

<sup>14</sup> J. von Neumann and O. Morgenstern, *Theory of Games and Economic Behavior* (Princeton Univ. Press, 1944), p. 4. See also O. Morgenstern, "Limits to the Use of Mathematics in Economics," in *Mathematics and the Social Sciences* (Amer. Acad. of Polit. and Soc. Sci. June, 1963), pp. 12-29.

world that the politician or journalist knows to be highly relevant to the solution of economic problems, too many economists (but certainly not all) seem to want to spend their time only on the fanciest kind of theoretical needlework. I suspect that the explanation for this, again, is to be found in the sociology of the economics profession and its priestly scale of values.

But this greatly reduces the ability of economists to counsel or educate the public. If economists wish to play their public role more effectively, then they will have to try to acquire something like the broad, but detailed, knowledge of public affairs and of the springs of human action of, say, a Walter Lippmann or a James Reston. These are high standards, but I suggest them as models, not minimum requirements. I might similarly have suggested that we seek to emulate such great teachers from our own field as David Hume, Adam Smith, J. M. Keynes, and Joseph Schumpeter, whose greatness consisted in being students of society and not merely of economics. If, in saying this, I give offense to some able technicians, I apologize; but my aim is to attack the evils of scientism and too constricting a professionalism, which I think are spoiling economics, but not to attack good and clean technical work.

#### IV

Thus far I have given two reasons for the public's economic "illiteracy": (1) economics is a difficult subject, in which simple, basic, and comprehensible truths are hard to come by; and (2) economists generally have been ineffectual or confusing teachers. But there is a third reason that I must mention, however briefly: the inadequacies of the public as students and its resistance to economic reasoning.

Much of the public, I am afraid, resists not only economic but any kind of reasoning. This lesson first struck me as a sometime lecturer to Army troops (can anyone who experienced it ever forget the horrors of "Orientation"?). The realization that most people are not very susceptible to reasoned arguments has been confirmed by my later experiences as an editor, civil servant, and local politician.

Why is this? In part, the public's resistance to reason is (to be as tautological as Wicksteed) irrational—lazy, dull, or nasty. It may reflect a deep streak of anti-intellectualism, simple inertia, or fear and hostility toward any social or economic innovation. (Alabama voter's complete letter to his congressman: "Don't want no changes." His congressman's complete reply: "Ain't gonna be none.")

Some resist current economic reasoning because of the economic ideas they have already borrowed, as Keynes said, from some defunct economist. The New York World's Fair will have a Hall of Free En-

terprise dedicated to the ideas of Herbert Spencer. Its sponsors declare: "The only right way to meet a depression is to let prices adjust themselves to the purchasing power. This is Nature's method but in the language of Herbert Spencer, 'When once you begin to interfere with the order of Nature, there is no knowing where the result will end.'" My mailbag brings me an endless stream of good advice—such as this:

The inflation of prices and deflation of values are a simultaneous indication of waste and crime. . . . Surely, if the mere mention of tax reduction is sufficient to spur business activity, just imagine what abolishing taxes entirely, borrowing, reducing expenditures and paying off 10 percent on the national debt annually would do. . . . Fifty years of chaos, confusion, stagnation, scarcity, poverty and war should be sufficient evidence to prove the demerits of the pseudo-economics [*sic*] for any nation. Herewith I respectfully submit for consideration a format of the Charter of Economic Order. . . .

This one was signed G. Szmak, Industrial Economist, and ended with "Copy to Hon. Earl Warren, Chief Justice, Cabinet Members and Congress, Economic Organizations and Publications."

Not everyone who resists "standard" economic reasoning is irrational or stupid. The high-cost U.S. clothespin or carpet or glass manufacturer who resists free-trade arguments probably knows what he is doing. So does the farmer who wants to keep price supports, the trade unionist who wants to push wages up faster than productivity and restrict access to his part of the labor market, or the professor who refuses to give up tenure.

People with vested interests, or vested morals, or just plain moral principles of their own may close the door to economists' conventional logic and produce logic of their own. One man will oppose anti-trust laws because they "interfere with the rights of the individual"; another will favor antitrust because it "protects the rights of the individual." Some will favor tax cuts, not because they understand modern fiscal policy, but because they want to pay lower taxes, or because they see lower taxes as a means of arresting the rise in government expenditures, and thereby of "preventing socialism"; while others will oppose tax cuts because they fear that they will overstimulate the economy, breed inflation, lead to government price and wage controls, and ultimately hurt profits and "bring on socialism."

It does no good to hoot and feel superior; one must be self-critical and aware that one's own reasoning or values may sometimes be ritualistic rationalizations of one's own special interests. One must try to understand what's on other people's minds and respond fairly and patiently. Those who resist the economists' reasoning are not necessarily scoundrels or fools. Many are sincere idealists (though idealism may sometimes be the foe of reason). Walter Heller was right, if indiscreet, in blaming much of the opposition to the tax cut upon public devotion to the Puritan ethic.

To be sure, there are plenty of fanatics and know-nothings in the United States—quite a few of whom (for reasons pure or purely pecuniary) spend their time promoting their own brand of economic education, and attacking yours. And if you get out on the firing line, they will attack you, *ad hominem*.

Yet, when the worst has been said, what it proves is that the job is not only difficult but extremely important. Just how important it is to increase this nation's social, economic, and political understanding, and to battle ignorance and fanaticism, was made terribly clear by President Kennedy's assassination and the events that surrounded it. As one Dallas businessman, Dawson Sterling, said a few days after the President's death, "Our responsibility, I believe, is to see that our fellow man is enlightened and informed. Only an educated, enlightened, reasoning, dedicated people can keep us free." This responsibility belongs also to the economist; it is as important as any responsibility he has.

## V

How can economists best discharge this responsibility to educate the public?

There are no easy answers. Certainly economists must devote more time and effort to addressing the public, less to each other. They must publish more and better books and articles for the public, lecture to the public, and help to advance the cause of economic education in the schools.

They must say what they mean, and say it clearly, arrestingly. They should take a lesson from our late poet, Robert Frost, who once remarked: "All there is to life is getting a meaning into a lot of material. . . . You've got to be sweeping and you've got to be pointed. You've got to come out somewhere, just as plain as a wisecrack or a joke." This may be tough for economists, with their abiding affection for convoluted argumentation and qualification. But they must try.

This is not merely a matter of style. Far more importantly, it is a matter of cleaning up economic doctrine, of deciding what is essential and what is verbiage, of shaping the content of economic thought into something that can be communicated usefully to young students, to high school teachers, and to the general public.

The task also involves—just as painfully—finding out much more about the process of education. What do we know of the art of teaching? What has been learned by those who have worked on the problem, and how can economists apply this to their own job? Answering such questions will require real effort, real humility—and openminded-

•

•

ness. Can television, for instance, provide a means of economizing on our best teaching resources? Although the results are not yet entirely clear, it appears that the national television course, "The American Economy," has been a useful device for conveying "a broad, general understanding of economics" to a great many people.<sup>15</sup> But what are the limitations of mass teaching of economics—whether through television, or newspapers and magazines, or books? How can the classroom teacher's time best be used to reinforce and go beyond what the mass media or books can do? We need better answers to all these questions—that is to say, more good people to work on them. And we certainly need more first-rate economists to work outside the colleges—in journalism, business, government, and other organizations. There is where some of the most significant educational work can and must be done. And trying to work within organizations will, I suspect, do economics itself no harm, and may do it much good. As Hughes has said, the professionals in organizations "will be—although in some sense bureaucrats—the innovators, the people who push back the frontiers of theoretical and practical knowledge related to their professions, who will invent new ways of bringing professional services to everyone, not merely to the solvent or sophisticated few. Indeed, I think it likely that the professional conscience, the superego, of many professions will be lodged in that segment of professionals who work in complicated settings, for they must, in order to survive, be sensitive to more problems and to a greater variety of points of view."<sup>16</sup>

Most vital of all, if economic education is to be advanced, it will require economists who care deeply about what they are doing. All economists, obviously, will not have the same positions on either analytical or policy matters. But everyone should strive to be honest about his own views and values, and fair to other's. This, again, is terribly difficult (maybe impossible) to do. But the economic educators' problem is much like that of the press. A good newspaper or magazine tries to report events fairly and objectively; but this does not mean that it reports blankly, stupidly, literally, superficially, or that it selects what it covers by random sample; it is not neuter, sterile, mechanical. Indeed, much of the American press was in fact betrayed by a false no-

<sup>15</sup> Thus far, evaluations of the national television course have been done by Dennis L. Nelson, "How Does the Achievement of High School Students Taking an Educational Television Course in Economics Compare with the Achievement of College Elementary Economics Students?" (master's thesis, Univ. of Minnesota, July, 1963); Campbell R. McConnell and John Richard Felton, "A Controlled Evaluation of 'The American Economy'" (Univ. of Nebraska, unpublished paper); and Phil Saunders, "A Preliminary Estimate of the Effectiveness of 'The American Economy' Television Course in Training Secondary School Teachers" (unpublished paper, available through Joint Council on Economic Education).

<sup>16</sup> Hughes, *op. cit.*, p. 666.

tion of "objectivity" into building up McCarthyism. The truth-seeker must go beyond a literal recording of the surface "facts," of merely "saying what the man said," while stopping short of asking why did he say it. A responsible journal must go on to ask what was the quality of his evidence, what other evidence was there to oppose his, what values shall we use or judgments shall we make, in deciding how to play this story—and how to editorialize about it. There is no scientific way to resolve these questions, but there is a great need for moral responsibility and involvement in the stream of history on the part of the press—and of economists—and of all others who aspire to be more than hacks. This involvement, I emphasize, must be in the service of truth, not at the expense of it.

I think the job of public education in economics, with all its difficulties, is far from hopeless. Economists have in fact done much to educate many important political leaders and members of what is now popularly or unpopularly called "The Establishment." We have had no genuinely asinine reactions to recession since the war—something that cannot be said about either the Great Depression or the 1937-38 recession. Part of the credit for this may belong to the cold war or other political, economic, and technological factors, but part surely belongs to the economics profession. At minimum, the increasing status of the profession, as evidenced by the establishment and role of the Council of Economic Advisers, as well as by the role of economists in other agencies, has blocked off Presidents and other policy-makers from crackpot advice and has made it far more difficult for congressmen to take their cue from cranks, living or dead.

A start—I say this hopefully—has even been made with the economic education of the general public, particularly with the help of our late President. The latest Harris poll, it seems, shows a swing toward support of the tax bill.<sup>17</sup> Whether economic logic has prevailed, or just the

TAX CUT TIMETABLE

	November	September
Delay until budget is balanced.....	39%	41%
Don't delay tax cut.....	46	36
Not sure.....	15	23

The shift looks significant, but it is worth noting that more than half would still delay tax reduction until the budget is balanced or were not sure.

<sup>17</sup> "Harris Poll: Support Grows for Tax Cut," *New York Post*, Dec. 2, 1963, p. 8. Mr. Louis Harris reported: "The main opposition to the tax cut, of course, is not that relief of the taxpayer is undesirable but rather that there should be no reduction until the federal budget is balanced. Back in the late summer and early fall, a plurality went along with this view. Now there has been a reversal that puts public opinion squarely behind a tax cut now without any further delay."

weight of the opinion-makers, is, however, far from clear. Economists cannot be content, in this democracy, to educate only the elite and to win a point once in a while on public policy. Economists have also a critical role to play in helping to bring into being a better-informed citizenry, competent to reason for itself and to act sensibly on important private and public matters.



## ECONOMIC ANALYSIS AND PUBLIC POLICY: THE GROWING HIATUS

By ROBERT R. NATHAN

*Robert R. Nathan Associates, Inc.*

The theme of efficiency runs through the entire program of this year's meeting of the American Economic Association. The title for this session refers to "The Product" of education in economics, which we can assume concerns all those who have taken courses in economics and especially those who have majored in economics as undergraduates or in graduate schools. It should be productive to study the performance of those who have been taught economics as well as to study efficiency in the techniques and processes of teaching. This paper will deal with the former. The latter would best be undertaken by specialists in education.

### *Background for Judgment*

The frequency with which one hears criticisms of economists must certainly bring mental anguish to economics professors and instructors and a determination to encourage the study of the effectiveness of their teaching of economics. This is not an easy task because the subject matter is highly controversial. There are few subjects on which almost every man and woman has stronger opinions and this is true of both the trained and the untrained in the field of economics. The degree of independence an investigator can bring to such an evaluation is suspect. Objectivity in such an appraisal is certainly desirable, but can the appraiser evaluate the "product" other than in the context of his own experience and interests and especially relative to his own views?

It is in the realm of formulation and implementation of economic policy where the most severe criticism of the role and performance of the economist is directed. The product of the teaching of economics—as distinct from business administration—is also more the object of brickbats than of medals and citations in the business world. In general, it is in the policy and applied economics fields from which most of the adverse criticism derives. And it is in these fields where the strongest conceptions and preconceptions and misconceptions prevail.

It will not be very helpful if the teachers of economics who resent adverse appraisals counter by citing the biases of the critics. It may be a valid defense but it is hardly conducive to constructive solutions to say that most criticisms come from those who do not like what economists conclude or propose. Criticisms derive from more than just

prejudices. No matter how much one discounts the clamor of liberals that the younger economists are too conservative, or the carping of conservatives that too many economists are liberal and socialistic, or the wailing of the policy-makers or operators that economists are too abstract and cannot relate their professional work to reality, one cannot avoid the fact that today there is great dissatisfaction with the product of the teaching of economics. This dissatisfaction is widely prevalent among economists as well as others.

In anticipation of charges of prejudice against the observations presented in this paper, let me at the outset plead guilty to having strong convictions. There are reflections of both bias and of obsolescence in this paper.

First, concerning bias, I regard myself as a liberal in the current popular—not classical—sense of the word. I am and have been primarily policy-oriented in the field of economics for most of the more than thirty years of my professional life. I am principally interested in the policies and the means by which the economy of this country and of other countries can be made to function more successfully toward the end that levels of production and standards of living increase and are maximized over time, consistent with a high degree of freedom.

Such a background and point of view explain a degree of impatience with what seems to be a trend toward less and less policy orientation among the young economists; with more concentration on mathematical tools of analysis and less on understanding the quantity and quality of empirical data; with decreasing recognition that the real world is somewhat at variance with the assumed conditions and relationships in so much theoretical analysis; with what appears to be less participation among economists in economic policy debates, especially those that might have overtones of political or group conflict; and with what looks like greater conformity and more complacency in our increasingly affluent profession. This catalogue of strong and critical impressions is not unique, at least in circles in which I travel and they are not small and confined circles.

The time has not yet come for me to enter a plea of superannuation, but obsolescence may well be suspected, especially because of my limited mathematical training. Nearly all economics courses thirty or more years ago did not require mathematical orientation. In fact, in those ancient days—economically speaking—the quantification of economic concepts was in a relatively early stage. The amount and quality of statistical data were hardly conducive to intensive and precise arithmetic formulations and refinements. If this type of obsolescence in these days of automated economics were to bar one from appraising

•

the product of economic teaching then a substantial portion of middle aged, let alone older, economists had better retreat to quiet corners for mutual consolation sessions.

These disclaimers are not put forward for the purpose of denying validity to what will be discussed but rather to emphasize that this subject is one which is difficult to approach entirely objectively and therefore the reader should be aware of the background against which this appraisal should be considered and weighed.

### *Environment for Judgment*

Just as it is helpful to look at the background of the one who seeks to evaluate the effectiveness of the product of economic teaching, so it is necessary to consider the broad environment against which such an appraisal is made. To judge economists on the basis of their contributions to economic policy and economic performance, one must review trends in the economy and the role of economists in relation to the performance of the economy. Those who have been taught economics do not have the authority to establish and to control the economic policies of this country or of any other country. But they can and should exercise considerable influence. Therefore, it is entirely fair to judge the effectiveness of the product of economic teaching against the performance of the economy. At least this seems fully proper in the areas of applied economics and economic policies.

Everyone will agree that the functioning of the American economy and those of Europe and Japan since World War II have been superior to past records and to what might have been expected. In relation to past fluctuations in the wake of major wars, the recent record has been commendable. In the immediate years after World War II there was serious inflation in the United States, but it was largely attributable to inadequate monetary and fiscal policies formulated and pursued during World War II. Insofar as economic stability is concerned, the four recessions in the postwar years were relatively short and moderate in severity. In recent years inflation has been far more feared than prevalent. The present period of recovery and relative prosperity has been quite extended and gives evidence of going on for some time. Yes, the post-World War II years have been affluent ones.

Various individual security measures originally introduced in the 1930's, such as unemployment insurance, old age pensions, assured rights of workers to organize and bargain collectively, bank deposit insurance, and the like—all served to moderate the impact of these recessions on individuals. The brevity and mildness of the postwar recessions have been attributed by many analysts far more to built-in stabilizers than to the discriminating use of discretionary fiscal and

•

monetary policies. Thus, one may question how much credit for this better record is due to our economic policies of recent vintage.

Even if it is concluded that the postwar record has been more favorable than in the past, is there basis for judging that it was satisfactory in the context of our times; that economists should be given a highly favorable rating for influencing sound and constructive and effective policies? Clearly the answer must be "no." In terms of rates of economic growth, the past several years have certainly been unsatisfactory. In terms of unemployment, the picture has been far from commendable. In terms of balance of payments, the deterioration in recent years was slow to evoke constructive suggestions for solutions. In terms of living standards and sharing in our affluence, large pockets of poverty have persisted and there are many and large distressed areas all over the country. In terms of adaptability to automation, the adjustments have been far from rapid and smooth. In terms of progress among the less advanced and poverty ridden nations, the income and production gap between the developed and the developing countries has grown. On the whole, the wasted resources and the lost potentials in the United States and the failure to achieve significant forward progress in the underdeveloped countries are hardly reasons for awarding honors to economic performance in the middle of the twentieth century. In fact, the gap between progress in the physical sciences and progress in the social sciences seems to be a growing one.

### *Theory and Practice*

No one should expect that even economics majors with advanced degrees will emerge from universities with full understanding of economic policy problems and prepared to make prompt and substantial contributions to their solution. This capacity they must develop through experience and continued study. But at least during their formal training they should have been fully exposed to the major economic policy issues facing this country and the world today and the curiosities of many of the young economists should have been aroused to the point of seeking to work and do research and to write in the field of policy.

As with all sciences, there is need for appropriate emphasis on pure science, but there is also need for emphasis on the application of scientific knowledge with imagination and care. What seems to have occurred in the field of economics is relatively greater and greater emphasis on pure theory and on research in theory and less and less on education in the areas of applied economics and in the important field of policy. It almost seems, despite the very substantial growth in the number of students and teachers of economics, that the proportion and possibly even

•

the absolute amount of time and energy devoted to applied and policy work in economics has diminished rather than expanded.

It would be interesting to undertake a study of the contents of the top professional economic journals over the past ten or fifteen years, classifying the contents in various categories ranging from pure theoretical treatment to policy or applied subject matter. The contents of each article would need to be studied in order to achieve a meaningful classification. It seems to me that the relative emphasis on policy and applied subjects has declined measurably.

Since accepting the invitation to present a paper on this subject I have had ample opportunity to talk with a great many economists and I have been surprised at the frequency of these same views. This does not prove these contentions, but many economists do believe that the recent product of economic teaching is not a very usable and practical product.

Even if it is true in limited degree that interest in work on policies has fallen behind, we must ask ourselves why this has happened. It cannot be attributable to a lack of need for policy work. Memories can be unreliable and it is usually a fact that the old days were never as good as they are recollected, but I would venture that the Great Depression evoked a degree of emphasis in the 1930's on applied economics and policy-oriented economic work which has no parallel in our present day. During the World War II period came another surge of interest and activity on the part of economists in applied economics and in economic policies. Our knowledge and understanding of the economic problems of the world in which we lived took a long step forward during the 1930's and 1940's. We have continued to learn about the functioning of our economy but at a seemingly slower pace. More seriously, we seem less concerned about the use and application of what we learn.

In part all this may have resulted from the fact that there was a hospitable environment for policy-oriented economists in government agencies and that there were not many alternative opportunities to governmental service for economists in the 1930's or even in the 1940's. Many universities did not increase their staffs and some even curtailed the size of the faculties in the 1930's. Also, the excitement and inspiration associated with fighting the depression in the 1930's and the war in the 1940's offered a stimulating and challenging climate which has not been duplicated. Yet, it is difficult to conclude that highly interesting and serious challenging problems do not exist today and that therefore the economists are not stimulated as they were twenty or thirty years ago. True, we do not have one-third or one-fourth of our labor force unemployed as we did in the 1930's. True, the country is not engaged in hot war as was characteristic in the 1940's. On the other hand, we

•

are still far from the millennium. In the context of the shrinking world, of the more sharply focused ideological competition, of the assuredly greater capacity to solve the problem of poverty, of the persistence of unemployment, of vast unsatisfied requirements for public facilities and public services, of great needs and even greater opportunities for accelerated economic development abroad, who can conclude that our best economists' talents and capabilities cannot be provoked and are not desperately needed. Yet we do not seem to be providing the talents and the focus and the emphasis and the results consistent with these opportunities, and the teaching of economics must bear part of the responsibility.

### *Challenging Economic Issues*

Let us look at some of the specific economic issues which face us, and then consider whether we are making the contributions toward their solution which one might expect from the growing number of trained economists, from the vastly improved storehouse and flow of statistical data, and from the seemingly unlimited amounts of money available for research grants, fellowships, scholarships, travels, lectures, and other supports. Judgments will vary, depending on the standards set and on what one regards as the range of responsibilities of those who have had courses in economics. Also, the environment within which the economist performs will affect his efficiency and effectiveness. In turn, the economics profession does influence the environment for economic policies or, at least, it should.

The lagging rate of growth and the persistence of wasteful levels of unemployment have aroused only limited concern and not enough constructive response from the economics profession, let alone from others who have studied economics. Probably more time and effort and written words have been devoted to whether there is a problem of growth and unemployment and just what the measurements are to the fourth decimal point, than to policies for speeding the pace of growth and reducing the level of unemployment. The fact that the annual average gap between our productive potential (at the unambitious rate of 4 percent unemployment) and our realized gross national product over the past six or eight years has likely exceeded the entire gross national product of the nearly half billion residents of India has hardly caused most economists to raise their eyes from their slide rules. There has not been enough concern expressed by economists to create the needed degree of public interest and debate on solutions which would bring results. Whatever interest has been manifested among economists has lagged behind public concern rather than led and stimulated public debate.

The level of debate on fiscal policy matters seems substantially less

intelligent today than it was ten years ago. In the business community, in the press, in the legislative branch of the government, and among the general public there seems to be less enlightenment with each passing year on issues of revenues and receipts, deficits and surpluses, and on the public debt. The problem is not simply a matter of whether government deficits are good or bad or whether the national debt is harmless or catastrophic or whether tax incidence is equitable or not. On these subjects there are always differences in views which will forever be as emotional as analytical, if not more so.

What is distressing is the low level of understanding and the lack of any growing perception and clarity of understanding of the issues involved. Overwhelmingly there is a tendency to equate government deficits with inflation. Tax reductions are welcomed as an expansionist force only if expenditures are reduced. Higher prices are considered as an inevitable consequence of higher wage rates. Increases in private debt and even in state and local government indebtedness are acceptable if not welcomed, but the rise in national debt is widely regarded as the road to some kind of bankruptcy. These are views of many who have had courses in economics as well as those who mouth clichés and know-nothing generalities. By now there are hundreds of thousands of Americans who have been taught some economics in formal courses, but their voices seem as muted or as misguided as the rest. Maybe distance lends enchantment, but it seems that in England and in many of the countries of Western Europe fiscal policies are dealt with at far more enlightened levels than here in the United States. They relate fiscal policies primarily to levels of production and employment and price stability rather than to deficits and surpluses and the national debt as such.

Where does one encounter public debates that deal forthrightly with the alternatives of increased public spending versus tax cuts? What has been proposed in the realm of fiscal policy as guidelines, not only for appropriate levels of aggregate demand, but also for the division between public and private uses of our output of goods and services? Where have the economists been in dealing with the problems faced by states and localities, in focusing on the economic issues associated with the reduction of federal taxes and the simultaneous increase in state and local taxes?

Most economists' prescriptions for solving the problem of unemployment relate to aggregate demand and therefore to the relation of government revenues to government expenditures, without, of course, resorting to that nasty eight letter word "deficits." The subject of income distribution became increasingly regarded as completely taboo. One is led to conclude that the way in which our income is distributed has no

significant effect on the functioning of the economy. It would seem that the income shares for the various factors of production or by income levels would be best ignored. The relationships between wages and prices and profits are discussed with respect to price stability or instability but seldom with respect to levels and patterns of consumption and savings and investment. Is it possible that this is too political an issue for economists or does it belong in sociology or political science or is it really unimportant from the economic point of view?

Less and less attention seems to have been given in public discussions year by year to the role of taxation in affecting income distribution by shares among the factors of production and by levels of personal income which, in turn, influence in varying degrees levels of consumption, savings, and investment. As more data have become available it has been possible to understand more precisely the incidence of different tax measures, but the policy implications seem to be increasingly neglected from the economic as well as the equity points of view. Economists ought to have much more to say about changes in taxation as they affect savings as well as their likely impact on incentives to invest. For example, liberalization of depreciation, the granting of investment credits, the reduction of corporate tax rates and substantial lowering of the top personal income tax rates will have an impact on savings as well as on investment and this really ought not be too complex a subject for serious attention in public debates on tax policy.

Might wage "guidelines" tend to bring a shift in income distribution away from wages and toward profits? Could this render more difficult the achievement and maintenance of full employment? Might the fear of cost-push inflation make more remote the prospects for greater growth and economic stability?

Is the achievement and maintenance of full employment feasible through exclusive reliance on aggregate taxation and aggregate expenditure policies of the federal government? Can an adequate expansionist fiscal policy be achieved through congressional legislation? If economists cannot hope to influence needed legislative policies, must they not play an important part in formulating and packaging economic policies in feasible terms as well as in ideal terms?

The critically important subject of our balance of payments has probably received more recent attention than almost any other single economic problem. The bulky volumes of submissions to the Joint Economic Committee contain many excellent analyses. However, even on this critical subject the degree of attention devoted to solutions has been somewhat less than one might have expected relative to the supply of economic talent available for prescriptive purposes.

The role of urban development has certainly not received the degree



of attention from the economic policy point of view which its magnitude and its potential impact on employment and on living standards would warrant.

The economics of education has received growing attention from some economists, attributable in part to the interest and efforts to separate the various factors influencing changes in productivity. The economists have contributed to a better understanding of the fact that education does serve to increase the productivity and income of our population, but there has been limited interest in relating the needs and opportunities in education to the financial requirements needed and financial resources available for a much expanded and improved educational system. The economic implications of federal, state, and local participation need far more attention than has been given to it.

In view of the pursuit of price stability and balance-of-payments difficulties, much more policy attention might have been expected on the subjects of economic concentration and administered prices. Can price stability be achieved by relying entirely and exclusively on voluntary restraint or can techniques be devised to prevent price inflation and still assure a fair degree of freedom in wage negotiation and in price determinations?

In the international field the emergence of regional common markets and the pursuit of freer trade pose many unsolved policy problems. The terms of trade have moved adversely for the raw material and commodity producing nations and some devices will be needed to help these countries. Strategies for dealing with expanding trade between the free and the Communist countries call for new imagination and new emphasis.

In the economic development area there persists grave deficiencies in the contribution of economists, though this is one field where increasing talents and time have been devoted to analysis and to policy work. In the May, 1963, publication of the Institute of Economics and Statistics, Oxford, *Bulletin*, Vol. 25, No. 2, Dudley Seers has written a most provocative article entitled, "The Limitations of the Special Case." This paper has also been published as Center Paper No. 28 of the Economic Growth Center of Yale University. Mr. Seers is not gentle in criticising the role of the economist in the area of economic development. He feels that economists are highly unrealistic and are doing relatively little toward solving the real and practical problems which exist in the underdeveloped countries. In one paragraph Mr. Seers states:

Since undergraduate teaching does not do much, if anything, to break down parochialism and to bring home the realities of the modern economic world, the recruit to graduate school easily falls into the trap of admiring technique for its own sake. If the content is

dull, the student turns to form, and this is indeed encouraged. Far from becoming more practical (as say, medical teaching does in its later stages), economics becomes still more abstract.

Mr. Seers quotes H. R. Bowen as follows:

Graduate students of economics are confronted with a vast array of concepts, techniques, and detailed theoretical constructions. They perforce live in a strange world of indifference maps, kinked demand curves, cross elasticity, marginal propensity to consume, liquidity preference, net national product, sampling error, linear programming, and input-output matrices. They spend much of their time gaining familiarity with specialized concepts and techniques, and their success as graduate students is gauged largely by the degree to which they master them.

Two general observations should be made following this array of observations on policy issues. First, the list of issues is not all inclusive. The degree of deficiency in performance is not the same for each problem or policy area. Second, is the recognition that there are important and substantial exceptions among economists to the generalizations prescribed herein. Many economists who might read this paper should conclude that these statements could not possibly refer to them. Many have devoted much of their talents to policy areas. The likelihood is that many of these economists also agree that the product of teaching of economics is not entirely satisfactory and that far too many economists devote little time or energy or interest to the critical policy problems which face the United States and other countries in this time of great opportunity to make unprecedented inroads into the problems of poverty, stagnation, slow growth, inflation, trade restrictions, inequities, and so many other major issues.

### *Economists as an Active Force*

It was not intended in listing policy issues to even hint that there can be or should be only one answer. None of these issues lend themselves to one view that reflects the ultimate truth. Most of these are controversial subjects. They can and should be debated. My complaint is that there is not enough debate of policy issues by those who have studied economics and much of the discussion that does take place bears almost no resemblance to logic, reality, or sound economics.

There does not seem to be enough interest and enough conviction among economists and others who have been exposed to the teaching of economics to zero in on many of these crucial issues toward the end that at least the policy debates will be meaningfully focused and the decisions will be predicated on a fair degree of understanding of their implications and of their consequences. Is this expecting too much? Is it fair to place any or much or all of the blame on economists?

There are many who sincerely believe that none of the blame should be attributed to economists or to those who teach economics because

it is their obligation to teach principles and concepts. They believe that economic policy decisions are made on political grounds and therefore the economist, being primarily a scientist and not a politician, should not be held responsible for economic policies and economic performance. They say that economists are scientists and that they should not be expected to propose nor fight for economic policy but that they should confine themselves to providing the conceptual framework and the analyses upon which policies can then be formulated and implemented by others.

It is not the view of this paper that the product of economic teaching should be held entirely responsible for economic policies and economic performances in our society. However, they should be held accountable in considerable degree. This assumes, of course, that if there is knowledge and if there is understanding of the impacts of alternative policies, the decisions will bring improved performance. This is not a certainty, but I do believe that if all those who study economics could acquire a better understanding of how our economic society performs and if they helped spread such understanding, the record of our economy's achievement would be much superior to what it has been. It is difficult to believe that the level of serious debate of economic issues could be as low as it is if those who have been taught some economics really understood some basic economic characteristics of our system, or if trained economists cared enough about policy matters to try to spread knowledge and understanding.

Perhaps the discussion has been focused too much on the role of the economist. Important attention must also be given to the teaching of economics to others than those who become professional economists. The total number of college graduates who are exposed to courses in economics is certainly very large relative to those who actually are or regard themselves as professional economists. These noneconomists who have taken economics courses include individuals in all kinds of professions and all walks of life. Even politicians have had courses in economics. Business managers have been exposed to the teaching of economics. The utter lack of any evidence of real understanding of economic principles or economic issues by most of these people is appalling. In the political world and the business world, among professional men and leaders of organizations and many groups, including labor, and among administrators, public and private, one would hope and expect serious attention to be given to those economic issues which affect our fortunes and our lives. But usually the level of understanding is so low as to be completely and devastatingly frustrating.

Business schools too often concentrate on vocational courses and give too little emphasis to economics. Yet the businessman ought to

•

know and understand much about the economic society in which he lives and participates and performs day by day. Is his interest stimulated by those who teach him his course or two in economics or does he find the teaching of economics so abstract and unrelated to real life as to look upon the subject with boredom or contempt? If the interest of economics teachers in policy areas is lacking, then it can be well understood that those who are exposed to one or two courses in economics will not be policy-oriented. Unfortunately, the lack of effectiveness of the product of economic teaching is characteristic of those who have had much exposure as well as those who have had little exposure.

### *Theory and Application*

I have the utmost respect for and appreciation of those who are engaged in the development of theory, in the elaboration of basic concepts, in the conduct of pure research, in the development of precise techniques of analysis, and of those who teach or do research work in areas which do not fall within the applied or policy realm. The need for more and more refinements, clarifications, and modifications of theory is ever present. For the long run, our understanding of economics will be greatly enhanced only if considerable energy and effort and brains are devoted to theory. Similarly, the tools of analysis need to be expanded and sharpened. Mathematics has contributed greatly to more efficient and more penetrating analytical approaches. The real issue derives from what appears to be a growing tendency to concentrate less and less economic work on the applied and policy areas.

Theory and application are both needed. There is the problem of relative emphasis in the use of available resources. If the teaching of economics and if the work of economists were to become less and less related to reality, we might emerge with a most advanced and scientific set of principles, but the real world in which we live might reveal a less and less satisfactory performance in economic terms and less and less respect for the economics profession. This is what seems to have been taking place. Progress in theory appears to have been associated with a regression in the level of understanding of economic problems and policies and in economic performance. At least there has been slippage relative to the resources available, if not in absolute terms.

We must again ask ourselves "why," assuming that there has been relatively less emphasis on applied and policy economics. There is no lack of need for policy work and applied work. Does the explanation lie in the limited opportunities for economists in government service as compared with twenty or thirty years ago? Does this limit the chances to gain practical experience in the policy field? Did the trend toward conformity, the fear of controversy and the anti-intellectualism of the post-

## DISCUSSION

JOSEPH A. KERSHAW: I find myself in substantial agreement with Mr. Silk's account of the profession's failure to communicate with the population in general. Indeed, in some ways he may be too easy on us. A friend of mine who is high in the councils of the Administration said to me after a stint of testifying on an economic bill, during which no inkling of understanding of the simplest economic motions was evident, "These are not men who have never gone to college; most of them have, and most have had courses in economics there." Economic illiteracy exists, not only because people do not study economics, but even when they do.

That economic illiteracy is widespread is I think not open to doubt. Mr. Silk uses the case of the tax cut proposal to illustrate this, and it is a persuasive case. There are of course many others. Who has not watched with dismay the annual madness when grown men soberly debate whether to raise the debt ceiling to accommodate a revenue and expenditure situation these same grown men have already voted for! And what about hearing our most senior labor leader refer to automation as an unmitigated evil? Sometimes one thinks that Lewis Carroll must have done most of our economics teaching.

The interesting question of course is why. Mr. Silk attributes it to the difficulty of economics, the ineptness of economists as teachers, and the inadequacy of the public as students. I hope the second is wrong, and the third sounds more like a restatement of the problem than a reason; the first is surely at the heart of the problem, and I would like to elaborate on it just a little. Ours is a difficult subject. That cutting taxes without cutting expenditures (at least not as much) will reduce the deficit is a subtle notion. It runs against all notions of thrift and arithmetic as well as common sense. To understand the reasoning requires a degree of sophistication we are sometimes unwilling to appreciate. The wonder is, perhaps, that the proposal has the support it has.

Also, there are lots of economic problems that we do not understand, even though they seem simple. What should an underdeveloped country do to get its growth rate up? We used to argue that if only it could get its saving rate up, it would be off to the races. Now some are beginning to wonder whether this is really the significant impediment. In general, we have not covered ourselves with glory in telling these countries how to grow.

Another thing that makes the economist's job of communication difficult is that many of the terms we use are also in common use, usually with different meanings. Physicists talk about ergs and molecules and nuclei and people react with awe and respect. But terms like savings and investment, for example, have no mystery about them. Everyone knows that the purchase of stocks is investment, and that saving is what is left over after you have paid your insurance premiums, made your mortgage payments, and met this month's refrigerator installment.

Finally, as in the tax cut example, a great deal of economics does seem to violate common sense and this, coupled with the propensity of economists to argue heatedly in public, leads many people simply to refuse to believe

what economists say. Mr. Silk points out that public disputation is still very much a part of our profession.

What can we do about all this? Mr. Silk is relatively hopeful that things are improving and will continue to do so. I think I share his modest hopes, though I have periods of despair. The profession, the C.E.D., the Joint Council on Economic Education, and others are working on the problem. The television program last year is bound to have had an effect at the secondary school level. There has recently been a test for economic literacy drawn up for use in secondary schools, by a committee of which I was privileged to be a member. This test will generate more concern, I think, and hence more effort to correct the situation.

In the realm of public affairs I agree with Mr. Silk that the Council of Economic Advisers—and in particular the quality of the appointments to it and its staff—has raised the status of the profession and has made sure that competent advice is at least offered, whether or not accepted. These developments are hopeful; I suppose the real issue is whether they will pay off before the illiterates can bring us to our knees. It may be close!

To turn briefly to Mr. Nathan's paper, which reached me too late for the careful study it deserves, one reads it I think with two emotions. One is admiration for the obvious sincerity with which his indictment is made. The other is the fervent hope that he is wrong, at least in part. His main concern I take to be that the younger economists tend not to be policy-minded; they seem uninterested in policy questions on grounds that suggesting policy changes is somewhat unscientific. They are more interested in inverting their matrices, deriving their shadow prices, and getting their linear programming equations to be consistent. The Greek letter is glorified; the Arabic number denigrated.

Mr. Nathan concedes that this may be a sign of age on his part, that it may therefore be taking place in his imagination rather than in the objective conditions of the times. I do not know which it is, but I do think we should be careful not to discourage the younger and mathematical people from working on their models. They should be urged to keep some end use in mind (and my experience is that they are by and large understanding on this score), but they are making some important methodological advances and can be expected to continue doing so, which seems all to the good.

Mr. Nathan returns to the question of economic illiteracy in general, as I think he should. The seeming indifference of the public and the Congress to our ability to come to grips with either the unemployment or the growth deficiencies can be put down to lack of understanding (and perhaps fear of government, which may be the same thing) and this brings it back to the profession's doorstep. Mr. Nathan and Mr. Silk meet on this, and for the very good reason that it is the heart of the problem.

I would like to close these brief comments with a paragraph or two on the subject I thought Mr. Silk and Mr. Nathan would write about. This is the problem of measuring the output of the teaching process, whether the subject be economics or any other. As economists, of course, we are interested in the efficiency of any process of production, and we ought to be doubly interested

in efficiency in education, since education is such a huge consumer of our resources. Small gains in efficiency yield large gains in resources saved.

Economists have not contributed much to the solution of this problem. The prime difficulty is that we have been unable to specify, let alone measure, the end product of the process. The difficulties are well known: how does one handle research at the university level, is learning of subjects the goal in high school, or are we interested in such things as social adjustment or creativity as well, and so on? Lacking a specifiable and quantifiable output, one cannot compare various combinations of inputs. This is not unlike the matter of making intelligent choices in the national security field, and one wonders why we seem to have done so much better with that range of decisions. It would be a gross exaggeration to say that notions like cost-benefit analysis have been embraced by either the military brass or the Congress, but the principal civilian decision-makers in the Defense Department are very much at home with them.

I would hope it would be possible to make similar progress in the analysis of educational efficiency. The measurement of the product ought to be possible, given the tests available today, and if the measurement is of only a partial output, this at least would be a start. With effective before and after achievement indicators, it should be possible to isolate extraneous factors, try different techniques, cost them, and then choose the best combination of inputs. Educationists do what purports to be this in profusion. But they do not understand the idea of incremental gains and costs, the interaction of inputs on each other, and how to handle extraneous factors that are overwhelmingly important, and they think anyone interested in dollar costs is out to cut salaries or to kill a bond issue. These are matters for economists who, as Mr. Silk reminds us, do have the conceptual apparatus rather specialized to these sorts of problems, and we ought to be getting at them.

KENYON A. KNOPF: As discussants at professional meetings will do, I have written my own paper on the subject, and it corresponds rather well with the paper I received written by Leonard Silk and only a bit less well with Robert Nathan's paper. That is not too surprising. Most critics of our performance as educators see approximately the same causes for our deplorable record. They also seem uniformly unable to get the profession very far along the road to reform. Hence, I am happy to accept most of what Mr. Silk and Mr. Nathan say with only slight change of emphasis, but I doubt that much will come of this latest effort to rouse the profession. Change appears to be a matter of small increments at the margin.

Leonard Silk spends little time to establish that the teaching of economics is inefficient when the test of efficiency is economic understanding by the public or public leaders. Mr. Silk does not need to support his case even when speaking to economists. We still are sick from the large doses of misunderstanding in the year-long debate over tax reduction.

What are the causes of inefficiency in the teaching of economics? Mr. Silk says there are three: economics is a tricky and difficult subject; economists generally are poor teachers; the public resists economic education. Putting these together we have this proposition: because economics is a tricky

and difficult subject, economists must be good teachers to overcome general public resistance. Mr. Silk illustrates the complexity of economics and economic jargon—important sources of public resistance. Mathematicians have been frustrated for some time by resistance to their subject from similar sources. They have thrown themselves into solving the problem—as a profession—and we can see the remarkable results in entering college students today.

Mathematicians have had to overcome fear; economists have to overcome suspicion, because economics not only is difficult, it is tricky. What Paul Samuelson has called “the fallacy of composition” produces overwhelming frustration. Things just are not what they seem. Common sense or personal experience often provides completely wrong answers. The economists’ world is a topsy turvey one to the layman; yet we blandly and routinely run through our explanations expecting students and the general public either to follow or to believe, despite the fact that they must ignore almost all points of reference familiar to them. A few common examples should suffice: (1) individual saving is usually a good thing for the individual, but if a large number of individuals try to save more, they may altogether save less; (2) a reduction of wages may expand employment in a single firm, but if a great many firms lower wages, employment may be adversely affected; (3) if one farmer expands output by working harder, he will increase his income, but if very many farmers do the same, their income will fall; (4) if one consumer annually increases his indebtedness, he will be in trouble, but if all consumers together expand their indebtedness, the economy and consumers will prosper; (5) an individual whose indebtedness grows regularly, faces bankruptcy but a society whose (government) debt grows may thereby assure its prosperity.

Public resistance to economic understanding, then, is substantially stiffened by the fact that a course of action that is bad when pursued by an individual often is good when pursued by large groups or by society. Alternatively, a course which is good for the individual may be bad if pursued by a large group or an entire society. Economists speak difficult conundrums. We are incensed and contemptuous when the public does not play the game or refuses to believe the unbelievable. We too easily accept special interest and stupidity as the only causes of resistance, forgetting how strange our wonderland is to others.

Why are we not more efficient in professing economics—in spreading economic understanding? Mr. Silk presents an excellent treatment of the problem of professionalism. Professionalism is necessary and laudable, but it creates a curtain between professional and layman which can be penetrated only by great effort on the part of the professional as well as the layman. Economists generally have not made the effort, perhaps because it would appear to produce a misallocation of resources in a market system characterized by consumer sovereignty. The general public, according to this view, has not “voted” for economics instruction and economists have disdained forcing their attentions on the public. Such action would be closely akin to advertising or “selling” and we all know how these activities misallocate

•



resources! Like polite children, we'll respond when asked. The President and Congress have asked and we have accepted the staff positions offered. Businesses and labor unions have asked and we have been employed preparing background papers for policy statements. Now there is a close call! But, of course, economists are technical experts, while businessmen soil their hands issuing policy statements to influence and instruct.

Too many economists fail to realize that the general public have voted for economics education; they just have not got it. We have an imperfect market here. The buyers are hundreds of thousands of college students who decide to spend some of their tuition dollars on undergraduate economics courses each year. Over a couple of decades, millions will have tried to buy economics instruction. The sellers, of course, are the economics professors in the colleges and universities of the country. Market imperfection appears when the sellers are not induced to produce what the buyers stand ready to buy. Rewards to sellers seldom are scaled to excellence in teaching; rather they are scaled to research and, more especially, to volume of technical publication. Economists, like businessmen, are quick to respond to the realities of their market situation.

The concept of professionalism is quite relevant to Robert Nathan's paper, too. This is the essence of his finding that the study of policy problems has declined relative to the study of theory. (I wonder if it really has declined. Hasn't it all just been lumped under the title of economic development?)

What can be done to correct the situation? We could reconstruct the market to put buyers in more direct contact with sellers, something in the manner of European institutions of an earlier era. But changes in production functions have made such reconstruction difficult for education as for business. Just as a business ethic seems necessary—an ethic which embraces a responsibility to society beyond that which is forced by the market—so is an economists' ethic needed, one which recognizes a responsibility beyond that enforced by the present system of rewards and punishment. We need a general response to the task Mr. Silk sets us, not a response limited to a few members of an association committee, or a growing yet small number of economists whose reputations are so well established that they can "afford" the time to care about economic education, or a somewhat larger group who seem not to care about furthering their professional reputations. We need to recognize a Robert Heilbroner as more than an outstanding journalist whose lucid books we recommend; such a remarkable interpreter must be recognized as making an important contribution as professional economist, equal to or more significant than much of what passes for research today.

What else can be done? Robert Nathan suggests more emphasis on policy. This might come about through some shift of emphasis in theoretical research. Gunnar Myrdal<sup>1</sup> has criticized the American economics profession, I think properly, for neglecting "the long range prospects and [for] the concentration on short-range issues and also on, by any standards, less impor-

<sup>1</sup>In an address to a convocation of the Fund for the Republic as reported in *The Des Moines Register*, Des Moines, Iowa, January 31, 1963, page 4. •

tant theoretical problems. . . ." He goes on to say, this "is probably responsible for the conspicuous failure of my distinguished and numerous colleagues to disseminate a reasonable degree of economic understanding to the American people, in spite of the fact that America has a larger college attendance than any other Western country."

Beyond this, as Mr. Silk points out, we generally have failed to define that dividing line between economic analysis about which economists generally agree, and judgments about the state of the economy and the efficacy of policies for that state, about which economists disagree. No wonder that people turn from economics with the idea either that one opinion is as good as another or that all economic positions are dogmatic. We must pick up arguments based upon fallacious propositions and destroy those propositions publicly in generally understood English, and then point out where the real area for disagreement lies. The authors of the principal papers do not advocate a weakening of research effort, nor do I. We do press upon economists their obligation to make research available to the general public through excellent writing, public debate, and, above all, excellent teaching. In short, this profession must determine to profess economics, not just to one another but to all who can be induced to listen.

HENRY H. VILLARD: I gather that Messrs. Lewis, Nathan, and Silk are not particularly happy about the level of economic literacy that presently prevails—which implies that there is something wrong with the product we produce. It seems to me that their conclusions are grossly unfair to economists—and that the time has come to defend our profession. Can any of you name any economist who accepted the achievement of general economic literacy as a major—or even an important—objective for economists? Or can you name any economist who has argued that we have an obligation to serve the general public, as Dr. Silk has suggested? I know of none—and I can see no reason why we should be criticized for not achieving what we never sought to achieve!

What have we sought to achieve? It is not a matter to which we devote much explicit attention, but Paul Samuelson's formulation, already quoted by Dr. Silk, seems to me to hit the mark: "The economic scholar works for the only coin worth having—our own applause." Actually I see no reason for including "economic" in the sentence. All scholars—including theologians debating the number of angels that can dance on the point of a needle—work primarily for the applause of other scholars. Would we really have it otherwise? Think of the serious psychoses that would develop among theologians if general comprehension of the behavior of angels were to be made the basis for approval of their activities.

For those of you who still have doubts or think I exaggerate, I can demonstrate that we have never been concerned with general economic literacy by considering what we would have done if we had had any such concern. First, if one wants to market a product, the thing to do is to research the market for the product. Even today less than 30 percent of Americans attend any type of college. With a handful of exceptions—Dr. Silk among them—

professional economists have been utterly uninterested in the level of economic comprehension which prevails in our high schools. How then can it be argued that we are seriously marketing economic literacy when 99 44/100 percent of economists are completely unconcerned with 70 percent of the market for our product?

Mind you, if we were interested, we would have to work at writing in clear English, capable of general comprehension. Is anyone prepared to insult our profession by arguing that what appears in our professional journals—loaded as they are with mathematics and models—represents a serious effort to write clear and comprehensible English? This is a point that Dr. Silk has covered. I have nothing to add except to reject as obvious slander the notion that we try to write with clarity.

What of the 28 percent who do attend college? Roughly three-quarters manage to avoid taking any economics, so that our product is currently reaching perhaps 7 percent of those who are currently being educated. Admittedly market penetration by *Rolls Royce* is significantly less. But *Rolls Royce* has a great advantage: it can charge five or six times as much as the bulk of its competition. We have no such easy way out, as it does not cost a student more to take a course in economics. So if they stay away, it would seem to be a fair reflection of the extent to which we have succeeded in making our product unattractive.

What do I mean by "unattractive"? Let me take an example that is close to home. Our wives, having had the good judgment to marry us in the first place, are usually inclined to show continued good judgment in accepting our opinions on economic matters. But how many of them could discuss President Kennedy's tax proposals in a way to satisfy Professor Lewis? I fear me economic illiteracy begins at home. But—and here is the point I want to make—how many of us would recommend Samuelson to our wives to provide them with insight on the pros and cons of a tax cut at the present time? Let me stress that I am not casting aspersions on the intelligence of our wives: I feel certain, as a group, they are much more intelligent than the average college graduate. My point is simply that, if we would not recommend Samuelson to our wives, does its incredibly widespread use have something to do with explaining why 75 percent of present college students carefully avoid taking any work in economics? (May I say, parenthetically, I am picking on Paul only because it seems a fair price for him to pay for outselling the competition in so decisive a fashion! My point really applies equally to all texts with widespread acceptance at this time.)

You may ask why, if my contention is correct, we continue to use Paul's book. I give you again Paul's own answer: as economic scholars—and who in this publish-or-perish world dares deny he is a scholar—we choose texts, not on the basis of student needs, but to earn the applause of other scholars. After all, a man teaching in East Keokuk Junior College has little enough to brag about when he gets to a convention; who would deny him the thrill he gets when he is able to say, "Of course we too use Samuelson"!

Again, if it is not to intrigue the instructor—who, after all, does select the text—why should we so complicate and elaborate the obvious? Paul uses

as the text for his chapter on "The Theory of Income Determination" a quotation from Keynes: "Given the propensity to consume and the rate of new investment, there will be only one level of employment consistent with equilibrium." Now recall that the reciprocal of the propensity to consume is the relationship between investment and income and that Keynes's assumptions in regard to wage units established a fixed relationship between income and employment. What, therefore, Keynes is saying is that, given the level of investment, given the relationship between investment and income, and given the relationship between income and employment, there is only one level of employment consistent with any given level of investment. This seems to me to be a relatively simple observation; in fact, not to put too fine a point on it, I am inclined to describe it as a tautology. But by the end of Paul's chapter—which is fourteen diagrams, seven 45 degree lines, and one table later—I find my students tend to lose sight of both the limitations and insights that stem from the Keynesian formulation.

Perhaps this is as good a place as any to say that there is another point in Professor Lewis' letter to Dr. Silk that bothers me. He seems to imply that what people say about the proposed tax cut should provide evidence that some "economics has been taught in the present century." Why so? For there is nothing in the economics we taught before 1948, when Paul's book first appeared, that sheds light on why taxes should be cut at this time. Certainly the sixteen pages devoted to the business cycle out of the thousand pages that made up Furniss, Fairchild, and Buck, from which I was taught thirty-three years ago, provided no insight. But, of those presently over twenty years of age, not many more than two out of seven have become twenty since 1948. From this I deduce that hardly more than 2 percent of the population—2/7's of the 7 percent who ever took a course in economics—have taken one since 1948. Even if we assume that everyone who has taken a course since 1948 has a firm grasp of Keynesian economics—which is quite an assumption—it still follows that hardly more than 2 percent of those over twenty should be expected to talk intelligently about the tax cut on the basis of what they were taught in college. Under the circumstances, I am mainly surprised that Professor Lewis should be surprised over the nonsense that is floating about!

Last, but not least, if we were serious about marketing our product, we would have to introduce some standards of relevance. As long as we are only talking to each other—seeking our own applause—what one economist talks about needs to relate, not to reality, but to the conversation of other economists; and we can define economics as what economists happen to be talking about. But this approach raises marketing problems. In 1930, economists were talking about economizing and economics was defined by Lionel Robins as the science of economizing. As a result, my introductory course devoted much energy to showing that economic rent was a surplus rather than a cost and that additional units of the variable would be applied until balance was achieved at the intensive and extensive margins. Meanwhile, the stock market had crashed, unemployment was steadily rising, and the 1.6

•

•

percent of Furniss, Fairchild, and Buck devoted to the business cycle provided no explanation of what was happening.

Is this sort of thing true only of the dim distant past? I do not think so. Admittedly economics has now been broadened to include macro- as well as microeconomics. But what of economic development? I estimate that whether we realize a 1.5 or a 3 percent annual increase in real income per man-hour will determine whether today's student retires with an income of \$13,000 or of \$23,000 (in 1960 prices). Under the circumstances, one might have supposed that introductory texts would concern themselves with how an economy might go about achieving 3 rather than 1.5 percent. But they do not—and I think the situation will get worse before it gets better. For I recently asked a young instructor, just completing his graduate work at one of our better graduate schools, why his proposed course outline omitted any discussion of economic development and confined itself entirely to micro- and macroeconomics. He replied, rather indignantly: "But that's all I was taught!" Mind, now that City has started to offer the Ph.D., I can see the problem graduate schools face: with no increase in time, they are expected to teach an increasingly complex body of theory—first in English and then in mathematics. The temptation to concentrate entirely on tools is immense; in fact, irresistible. And the inevitable result is that our introductory courses are increasingly crammed with tools and increasingly designed primarily for the benefit of potential majors. After all, how else can one explain why there is general agreement that the best way to review for the Ph.D. orals is to read Samuelson!

In sharp contrast, if we really wanted to introduce sales appeal into our introductory courses, we would have to make them relevant to the needs of our students—which means we would have to include material not taught in graduate schools and, even more important and more difficult, omit a lot of the material that is taught. All in all, to bridge the gap between what economists talk about and what students would like to know about is a tough assignment. I am not at all sure we should make the attempt: we might only succeed in inducing wholesale schizophrenia. But there is one thing about which I am absolutely certain: thus far we have not tried!

AMERICAN ECONOMIC ASSOCIATION

PROCEEDINGS  
OF THE  
SEVENTY-SIXTH  
ANNUAL  
MEETING

BOSTON, MASSACHUSETTS  
DECEMBER 27-29, 1963

## PROCEEDINGS OF THE AMERICAN ECONOMIC ASSOCIATION

### ANNUAL BUSINESS MEETING, DECEMBER 29, 1963 STATLER HILTON HOTEL, BOSTON, MASSACHUSETTS

The Seventy-sixth Annual Business Meeting of the American Economic Association was called to order in Parlor B of the Statler Hilton Hotel, Boston, Massachusetts, at 5:00 P.M. by President Gottfried Haberler. The minutes of the business meeting of December 29, 1962, were approved and the minutes of the Executive Committee meetings and reports of officers and committees of the Association were ratified. These reports are published in the "Proceedings" and constitute the official actions of the Association when approved at the annual meeting.

The report of the Secretary was presented by H. F. Williamson. The schedule of times and places for future annual meetings through 1970 was outlined. Attention was called to the fact that the total number of members and subscribers as of November 30, 1963, was 16,279, a net increase of 635 for the year. The Secretary reported that the 1964 edition of the *Handbook* was completed and would be distributed to the membership early in January; that the Johnson Reprint Corporation is now prepared to supply back issues of the *American Economic Review* for the years 1911-56; and that the total of subscriptions to the *Journal of Economic Abstracts* had reached nearly 7,000. He also called attention to two new activities undertaken by the Association: (1) the development of a roster of economists interested in overseas assignments and (2) the development of a national register of scientific personnel in cooperation with the National Science Foundation. In conclusion, he called attention to the fact that one group flight had been arranged from Chicago to Boston for the meeting and that similar arrangements are planned for future meetings.

Copies of the preliminary Report of the Treasurer were distributed by H. F. Williamson. Special attention was called to the fact that despite a reported outlay of over \$36,000 for the publication of the 1964 *Handbook* (a nonrecurring item), the net operating loss for the fiscal year would probably be relatively small. The full details of the financial operations of the Association for 1963 are shown in the tables in the Treasurer's report to the Executive Committee.

J. W. Bell reported for the Finance Committee on the changes in the Association's investment portfolio. The full report of the Finance Committee is published below.

The report of the Managing Editor, published below, was summarized by J. G. Gurley. He expressed his appreciation for the effective work done by the Editorial Board and consultants in preparing the *American Economic Review*.

The Secretary then presented the report of the Committee on Elections and the certification of the election of new officers for the year 1964, as follows:

In accordance with the bylaws on election procedures, I hereby certify the results of the recent balloting and present the reports of the Nominating Committee and the Committee on Elections.

The Nominating Committee, consisting of Paul A. Samuelson, Chairman, Robert E. Baldwin, Harry C. Eastman, Edgar O. Edwards, Frank W. Fetter, and Ilse Mintz, presented to the Secretary the list of nominees for the respective offices:

*For President-elect*

Joseph J. Spengler

*For Vice-Presidents*

James S. Earley  
Bernard F. Haley  
Oskar Morgenstern  
James Tobin

*For Executive Committee*

Joseph D. Coppock  
Harry G. Johnson  
Douglass C. North  
Robert M. Solow

The Committee on Elections, consisting of Richard B. Heflebower, Chairman, John H. Wills, and Harold F. Williamson, prepared biographical sketches of the candidates and ballots were distributed early in November. The canvass of ballots was made on December 20, 1962, and the results were filed with the Secretary.

From the report of the Committee on Elections, I have the following information:

Number of envelopes without names for identification .....	136
Number received too late .....	14
Number of defective ballots .....	-
Number of legal ballots .....	4,003
Number of returns from the mail ballot .....	4,153

On the basis of the canvass of the votes cast, I certify that the following persons have been duly elected to the respective offices:

*President-elect* (for a term of one year)

Joseph J. Spengler

*Vice-Presidents* (for a term of one year)

Bernard F. Haley  
James Tobin

*Members of the Executive Committee* (for a term of three years)

Harry G. Johnson  
Robert M. Solow

Following the reports of the Committee on Elections, President George J. Stigler took over the Chair and introduced Joseph J. Spengler, the new President-elect.

The President recognized Harry G. Johnson, Chairman of the Committee on Resolutions, who presented the following resolution:

Be it resolved that the Association expresses its appreciation and gratitude to all those who have shared in the planning, organization, and conducting of the Seventy-sixth Annual Meeting. The Association is particularly indebted to its President, Gottfried Haberler, for the preparation of a Presidential Address of deep scholarship, broad historical sweep, and great contemporary relevance; to its President-elect, George J. Stigler, for arranging a program centered on problems of fundamental importance to the membership; to the Executive Committee for assisting in the exacting task of planning and staffing the program; and to the many speakers and discussants who have presented and discussed new work in economics for our edification and enlightenment.

HARRY G. JOHNSON, *Chairman*  
W. ALLEN WALLIS  
DONALD C. RILEY

On behalf of the Association, the President expressed the appreciation of the excellent job done by the Committee on Local Arrangements.

There being no unfinished nor new business, the meeting was adjourned at 5:45 P.M.

HAROLD F. WILLIAMSON, *Secretary*



## THE JOHN BATES CLARK AWARD

CITATION ON THE OCCASION OF THE PRESENTATION OF THE MEDAL  
BY EDWARD S. MASON TO HENDRIK S. HOUTHAKKER

DECEMBER 27, 1963

Every other year the American Economic Association awards the John Bates Clark Medal "to that economist under the age of forty who is adjudged to have made a significant contribution to economic thought and knowledge." The award is made by a joint vote of the Association's Executive Committee and its Committee on Honors and Awards. This year the recipient of the medal is Professor Hendrik S. Houthakker of Harvard University.

Outstanding among Professor Houthakker's many contributions, both theoretical and empirical, are those he has made on the subject of consumers' expenditures. A particularly noteworthy characteristic of this work is the way in which the theoretical and empirical aspects are intertwined. As he himself has written:

If the theory of consumer's choice is to be more closely integrated with empirical demand research . . . it needs to be both generalized and specialized. Generalization is needed to deal with phenomena . . . that are outside the scope of the classical assumption; specialization is needed to give more substance, and hence more verifiability, to the very general theories that have traditionally been the theorist's stock-in-trade, and to enable the formulation of additional theories based on less general, but perhaps equally realistic, hypotheses. These . . . in turn provide new impetus and direction to empirical analysis.

Professor Houthakker has contributed significantly to the generalization of demand theory and to its extension to new realms in his work on revealed preference, rationing, and quality variations. He has contributed to the specialization of the theory, and thereby added to its significance for empirical research, in his papers on indirect utility, direct additivity, and systems of demand functions. In addition to these important and extensive contributions to the subject of consumer demand, he has brought econometric methods to bear on a number of important policy issues and problems.

On behalf of the American Economic Association, I have the privilege and the pleasure of presenting the John Bates Clark Medal to Hendrik S. Houthakker.

## MINUTES OF THE EXECUTIVE COMMITTEE MEETINGS

1. Minutes of the spring meeting held in New York City, March 29-30, 1963:

The *second meeting* of the 1963 *Executive Committee* was called to order at 9:30 A.M. at the Waldorf-Astoria Hotel, New York City, March 29, 1963. The following were present: Gottfried Haberler, presiding, Gardner Ackley, W. J. Baumol, Ewan Clague, E. D. Domar, J. G. Gurley, C. P. Kindleberger, A. P. Lerner, P. A. Samuelson, G. J. Stigler, W. A. Wallis, and H. F. Williamson. Absent were: E. S. Mason and Robert Triffin. Present as members of the Nominating Committee were R. E. Baldwin, H. C. Eastman, E. O. Edwards, F. W. Fetter, and Ilse Mintz. Present as guests were: G. L. Bach, B. F. Haley, and B. W. Lewis.

1. *President's Remarks* (Gottfried Haberler). President Haberler outlined the items on the agenda to be covered at the meetings.

2. *Minutes*. The minutes of the Executive Committee meetings of December 26 and 29, 1962, previously sent out to members, were approved.

3. *Report of the Secretary* (H. F. Williamson).

*Annual Meetings*. The Secretary reported that plans were well under way for the 1963 annual meeting at Boston. George H. Ellis, President of the Federal Reserve Bank of Boston, is Chairman of the Committee on Local Arrangements and Richard de Costa, also from the Bank, is Executive Secretary. The headquarters hotel will be the Statler Hilton. Attention was called to a suggestion by Donald Riley, of the American Statistical Association, that our 1970 annual meeting be held in Detroit. The Secretary was authorized to explore this possibility and report back to the Executive Committee at the next meeting.

*Membership Analysis*. The Secretary reported the total of members and subscribers as approximately 15,980. He pointed out, however, that until the Association shifted from the data processing company to its own system it would be impossible to give an exact count.

### *Publications.*

*Papers and Proceedings*. It was announced that the plan was to print 17,700 copies of the 1963 *Papers and Proceedings* and that the total manuscript pages would be approximately 1,221.

*Handbook*. The Secretary reported that plans for the publication of the 1964 edition of the *Handbook* were going ahead on schedule. Attention was called to a recommendation by Professor M. A. Copeland that the 1964 *Handbook* include two new lists: one, a division of members by age group, possibly putting those born during and before 1928 in one group and those born after in another; and, two, a special geographic listing for the largest cities under subheadings of academic members, business members, and government members, other members, and subscribers. After a discussion of various ways by which the *Handbook* might be improved, it was VOTED that the Secretary should explore the possibility of listing of members by subdivisions of special fields and that the *Handbook* should include a listing of members by their academic affiliation.

*Advertising*. The Secretary reported that the paid advertisements for the March, 1963, issue of the *American Economic Review* were 42½, up two and a half pages from the March, 1962, issue. Since the advertising rates had not been raised in several years, the Secretary was asked to review the rates and to raise them to levels consistent with the rates charged by similar publications.

*Association Archives*. The Secretary reported that plans were under way to set up an Association Archive during the summer months.

*Data Processing*. After a review of the difficulties the Association had experienced in respect to mailing lists and accounting records under the contract with the data processing firm, it was VOTED that the Association should return to the former card and Elliott stencil system as soon as feasible.

4. *Report of the Treasurer*.

*Bank Balance*. The Treasurer reported that the bank balance as of March 22, 1963, was \$21,591.00.

*Auditor's Report*. The Treasurer outlined the difficulty in getting an accurate report on the prepaid subscriptions and membership dues required by the auditors. It was

agreed that the Treasurer should discuss this question further with the auditors, with the view of obtaining data satisfactory to the auditors.

5. *Report of the Managing Editor* (J. G. Gurley). The Editor asked the Executive Committee to approve the list of nominees to be added to the Editorial Board. It was VOTED to approve the list.

6. *Reports of Standing and Special Committees.*

*Committee on Research and Publications* (W. L. Thorp, Acting). In the absence of the Acting Chairman, the Secretary read Mr. Thorp's report of the meeting of the Committee held in New York, February 26, 1963. The Committee: (1) unanimously recommended that the Executive Committee authorize the publication of Volume VI of the *Index of Economic Journals*; (2) decided to ask Bert Hoselitz to assume responsibility for the foreign economic research translation series; (3) agreed that it would be desirable to add volumes to the "Readings Series," giving priority to the fields of price theory and international trade, plus the possibility of a volume on macroeconomic theory.

*Surveys of Foreign Economic Research* (G. H. Hildebrand). In the absence of the Chairman, the Secretary read the written report submitted by the Chairman, which called attention to the chronic problem of the Committee in getting contributors to meet their deadlines. In view of these delays, the Committee decided that they should follow a policy of a single final extension in all cases for summer, 1963, even though this involves some risk of eventual delinquencies. It was also noted that the Committee was considering the alternative ways of publishing the surveys: either in groups of two or three as a separate publication or bringing them out seriatim, possibly as supplements to the *A.E.R.* Despite the difficulties encountered, the Committee hoped that by the end of the summer, 1963, most, if not all, of the essays promised would be delivered.

*Foreign Economic Research Translations* (Bert Hoselitz). In the absence of the Chairman, a memorandum prepared by Mr. Hoselitz was distributed, outlining the preliminary plans for the translation series. The members of the Executive Committee were asked to comment on this plan and to make any further suggestions as to possible works that should be translated.

*Journal of Economic Abstracts* (C. B. Warden, Jr.). In the absence of the Editor, the Secretary read the written report from Mr. Warden, which indicated that since December, 1962, Volume I, Number 1, of the *Journal of Economic Abstracts* had been published, and of the 28,500 copies printed, 7,850 were shipped out of the country for foreign distribution by contributing editors and 20,650 prepared in the United States for mailing to United States and foreign addresses. The publication of Volume I, Number 2, was scheduled for early April, 1963. The total number of subscriptions as of March 28, 1963, was about 3,500, with more coming in daily. The bulk of the subscriptions were from members of the American Economic Association and to a far lesser extent the Canadian Political Science Association. It was assumed that the delay in getting more subscribers from abroad could be accounted for by mailing delays for European and Asian addressees.

*Index of Economic Journals.* The Secretary read from a letter from J. P. Miller, strongly urging the Executive Committee to give serious consideration to the possibility of bringing out Volume VI of the *Index of Economic Journals*. After discussion, it was VOTED to ask Mr. Miller to undertake the supervision of the publication of Volume VI.

*Committee on Economic Education* (B. W. Lewis) and *Task Force on Economic Education* (G. L. Bach). The chairmen reviewed the activities of the two committees since their appointment, noting that the Task Force had completed its original assignment. Because of the growing interest in the general field of economic education in the schools, however, it was recommended that the President appoint a special committee to consider the future role of the Association in this field and to make recommendations to the Executive Committee at its meeting in December, 1963. Following a discussion, it was VOTED that such a committee should be appointed. The President announced that the members of the Committee would be B. W. Lewis, Chairman, G. L. Bach, R. A. Gordon, A. P. Lerner, W. A. Wallis, and H. F. Williamson.

*Committee on Honors and Awards* (B. F. Haley). The Chairman distributed a memorandum, outlining the problem of according greater recognition to distinguished members of the profession than is possible under the existing system of honors and awards. Four ways that such an objective might be achieved were suggested: (1) The establishment of the office of honorary president. (2) The establishment of a category of fellows or an expansion in the number of honorary members. (3) More frequent awards of the Walker medal. (4) An arrangement for the awarding of a new medal every other year. The Chairman noted that none of these suggestions had the full approval of the members of his Committee and that only in respect to the possibility of a group of fellows would

a bare majority of the Committee approve. Following a discussion, it was VOTED not to change the existing system of honors and awards. It was also VOTED that the Committee on Honors and Awards be authorized to decide whether or not to recommend a nominee for the John Bates Clark medal for any particular year.

*Nominating Committee* (P. A. Samuelson). The Executive and Nominating Committees met as an Electoral College to consider nominees for the office of President-elect for 1964. After discussion, the nominee was selected and his acceptance obtained. Nominations for the other offices were discussed.

*7. Reports of Representatives and Members of Advisory Committees.*

*International Economic Association* (B. F. Haley). No report.

*Institute of International Education* (Theodore Morgan). The Secretary read a letter from Mr. Morgan, indicating that the Ford Foundation had not, as yet, acted on the proposal for the continuation of the Economics Institutes. It appeared, however, that the Ford Foundation would provide funds for this purpose.

*Census Advisory Committee* (Solomon Fabricant). In the absence of the Chairman, the Secretary read a letter from Mr. Fabricant, noting the appointment of the following individuals to the Committee for a term ending December 31, 1966: S. J. Maisel, R. W. Pfouts, and Gideon Rosenbluth.

*National Science Foundation*. The President announced the appointment of O. H. Brownlee as the representative of the Association to the National Science Foundation for a term ending December 31, 1966.

*American Council of Learned Societies*. The President announced the appointment of O. H. Taylor as representative of the Association to the A.C.L.S. for a term ending December 31, 1966.

*Social Science Research Council* (G. H. Hildebrand). No report.

*National Bureau of Economic Research* (W. L. Thorp). No Report.

*American Association for the Advancement of Science*. The President announced the appointment of Bert Hoselitz as representative of the Association to the A.A.A.S. for a term ending December 31, 1965.

*8. Unfinished and Miscellaneous Business.*

The Secretary reported upon the suggestion of Henry Villard, of the Ford Foundation, that the Association establish and maintain a roster of American economists available for overseas teaching assignments. It was VOTED that the President be authorized to apply to the Ford Foundation for a grant to conduct a clearinghouse project for a three-year experimental period.

The Secretary reported that the Association had been asked by the National Register of Scientific and Technical Personnel of the National Science Foundation to establish and maintain a roster of members of the economics profession in the United States. It was VOTED to authorize the Secretary to make this application.

The Secretary indicated that he was exploring the possibilities of group flights for members attending the Boston meeting.

*9. Program for the 1963 Meeting* (G. J. Stigler). Mr. Stigler outlined his plans for the 1963 program.

The meeting adjourned at 11:00 A.M., March 30, 1963.

*2. Minutes of the meetings held in Boston, Massachusetts, December 26 and 29, 1963:*

The *third meeting of the 1963 Executive Committee* was called to order on December 26 at 7:30 P.M. at the Statler Hilton Hotel, Gottfried Haberler presiding. Others present were: W. J. Baumol, Ewan Clague, E. D. Domar, J. G. Gurley, C. P. Kindleberger, A. P. Lerner, E. S. Mason, G. J. Stigler, Robert Triffin, W. A. Wallis, and H. F. Williamson. Absent were: Gardner Ackley and P. A. Samuelson. Present as guests were: J. W. Bell, I. B. Kravis, B. W. Lewis, and J. J. Spengler. The meeting was adjourned at 10:30 P.M.

The *first meeting of the 1964 Executive Committee* was called to order at 7:00 P.M. on December 29, G. J. Stigler, presiding. Others present were: W. J. Baumol, E. D. Domar, J. G. Gurley, Gottfried Haberler, H. G. Johnson, J. J. Spengler, Robert Triffin, W. A. Wallis, and H. F. Williamson. Absent were: B. F. Haley, E. S. Mason, R. M. Solow, and James Tobin. Present as guests were: J. W. Bell, A. P. Lerner, and Arthur Smithies. The meeting was adjourned at 10:15 P.M.

*1. President's Remarks* (Gottfried Haberler and G. J. Stigler). Gottfried Haberler reviewed the order of items on the agenda, calling attention to unfinished business which might call for action before year-end. These matters were taken up first. The minutes, however, follow the order of business as presented in the agenda.

*2. Minutes.* The minutes of the March 29-30, 1963, meeting were approved.

3. *Report of the Secretary* (H. F. Williamson). The Secretary presented the official figures for registration at the 1962 annual meeting at Pittsburgh, the schedule of future annual meetings, membership growth and composition, publication costs of the 1963 *Papers and Proceedings*, requests to reprint, the use of the mailing list, and new activities of the Association—all more fully treated in the Secretary's Report. It was VOTED to approve and accept the Secretary's Report as presented.

It was VOTED to hold the annual meeting of the Association in Chicago in 1968 and in Detroit in 1970.

It was VOTED to hold the spring meeting of the Executive Committee in New York City, March 6-7.

4. *Report of the Treasurer* (H. F. Williamson). The preliminary report of the Treasurer was summarized. Special attention was called to the fact that because of the illness of a staff member, the final report was not ready for presentation at that time. As indicated by the preliminary report, the Association ended the year with a deficit of under \$4,000, even though there was a nonrecurring expenditure of approximately \$36,000 for the publication of the *Handbook*. It was VOTED to accept the report of the Treasurer, subject to final confirmation by the auditors.

5. *Report of the Auditor* (H. F. Williamson). It was VOTED to consider the Auditor's Report (which is presented below) at the March meeting of the Executive Committee.

The Secretary was instructed to write a letter of appreciation to David Himmelblau & Company.

6. *Report of the Finance Committee* (J. W. Bell). J. W. Bell outlined the main features of the report of the Finance Committee, published below. It was VOTED to accept the report and to reappoint the members of this Committee to serve during 1964. The Secretary was instructed to express to the members of the Committee the appreciation of the Executive Committee for their services.

7. *Report of the Managing Editor* (J. G. Gurley). The Managing Editor reviewed the report published below. He called special attention to the fact that the 1963 volume, like the 1962 volume, was larger than the 1961 because of the inclusion of the survey articles. He presented his proposed budget for 1964. It was VOTED to accept the report of the Managing Editor.

It was VOTED that Mrs. J. G. Gurley be given a vote of thanks for her services as book review editor without pay during the past year.

#### 8. *Reports of Standing and Special Committees.*

*Committee on Research and Publications* (I. B. Kravis). The Chairman commented briefly on his written report, published below, calling special attention to the new volumes to be added to the "Readings Series" and the progress being made on the translations of work of foreign economists. The Secretary noted that the preparation of Volume VI, *Index of Economic Journals*, was progressing satisfactorily under the direction of J. P. Miller.

*Surveys of Foreign Economic Research* (G. H. Hildebrand). In the absence of the Chairman, the Secretary reported that plans were now completed to publish the survey articles as supplements to the *American Economic Review* during 1964.

*Journal of Economic Abstracts* (Arthur Smithies). A written report, published below, was distributed by the Chairman. He called special attention to the fact that the total number of subscribers as of December 15, 1963, was over 6,000 and would probably number 7,000 early in 1964, and that Volume II, Number 1, of the *Journal* would be ready for distribution in January. The Chairman then raised a number of questions pertaining to the publication of the *Journal*, including: the advisability of adding abstracts from Iron Curtain periodicals; the feasibility of providing a classified index; the possibility of abstracting articles from the papers and proceedings of the annual meetings of various economic associations; and the desirability of publishing selected abstracts of articles from specialized journals. Following an extensive discussion of the foregoing suggestions, it was VOTED that the Editor be authorized to expand the *Journal* coverage to include selected abstracts from specialized journals. It was further agreed that the Editor should explore the possibility of including abstracts from Iron Curtain journals, as well as the question of a classified index. It was VOTED to authorize the Editor to accept advertising for the *Journal*.

*Committee on the Future Role of the A.E.A. in Economic Education* (B. W. Lewis). The Chairman presented a written report, published below. Following a discussion, it was VOTED to authorize the President to appoint a new six-man committee to explore the possibilities of improving the quality of economic education at the precollege, college, and adult education levels. Three of the members of this committee are to serve as an advisory committee to the Joint Council on Economic Education. Because of the appointment of this new committee, it was VOTED to dissolve the former Committee on Economic Education.

*Committee on Honors and Awards* (B. F. Haley). No report was presented.

*Committee on the International Economic Association* (H. S. Ellis). In the absence of H. S. Ellis, a written report of the Committee, published below, was circulated. Following a discussion, it was VOTED to appoint Gottfried Haberler to represent the A.E.A. in exploring the possibilities of obtaining financial support for future operations of the I.E.A.

*Nominating Committee*. The President announced the appointment of Gottfried Haberler to serve as Chairman of the Nominating Committee for 1964.

9. *Reports of A.E.A. Representatives or Advisory Committees.*

*A.C.L.S.* (O. H. Taylor). No report.

*S.S.R.C.* (G. H. Hildebrand). No report.

*N.B.E.R.* (W. L. Thorp). Report printed below.

*A.A.A.S.* (Bert Hoselitz). No report.

*I.E.A.* (B. F. Haley). No report.

*I.I.E.* (Theodore Morgan). The Chairman presented a report, printed below, and outlined the activities of the Institute during 1963, calling special attention to the fact that the Ford Foundation had made a grant of \$250,000, which it was estimated would enable the Institute to continue operations at least through 1968 and possibly through 1970. It was agreed that the Committee should make every effort to insure the future operations of the Institute without additional foundation support.

*Census Advisory Committee* (Solomon Fabricant). In the absence of the Chairman, a written report was presented, published below. In connection with the report, the President noted that Mr. Fabricant had raised the question as to the advisability of reappointing members of the Census Advisory Committee—more specifically, those whose terms expire at the end of 1963. Following a discussion of the desirability of assuring wide representation of the A.E.A. membership on the Committee, it was VOTED that members of the Census Advisory Committee could be reappointed at the discretion of the President but that no one should be reappointed after serving two consecutive terms until at least one year had elapsed. The Secretary was instructed to advise the Executive Committee at the March meeting as to the advisability of applying the foregoing policy to other committees of the Association.

*UNESCO* (K. E. Boulding). No report.

*Joint Council on Economic Education* (B. W. Lewis). No report.

*J.C.E.E. Special Publications Advisory Committee* (G. L. Bach). No report.

*National Science Foundation Board* (O. H. Brownlee). No report.

*National Academy of Sciences-National Research Council* (A. G. Papandreou). No report.

10. *Unfinished and Miscellaneous Business.*

The Secretary called attention to the fact that the cost of holding a special meeting of the Committee on Economic Education resulted in a deficit of \$181.49 in the Committee account. It was VOTED to approve this expenditure.

Theodore Morgan presented a draft proposal, worked out by a committee appointed by the Ford Foundation, of a program to improve the selection of students from foreign countries coming to the United States for graduate training in economics and agricultural economics. Following a discussion, it was VOTED to authorize the President to appoint a committee to consider the proposal and to report on its findings to the Executive Committee in March.

A letter was read from Professor Jacques Le Goff, of the International Economic History Association, inviting the American Economic Association to become a member and requesting a payment of \$25 membership fee. It was VOTED to accept the invitation.

The Secretary reported on a suggestion from the U.S. Employment Service to conduct the employment register on a twelve-month basis, with the central clearing operations to be handled in the Chicago office of the Illinois State Employment Service. The Secretary was instructed to explore this possibility.

The policy of the Association with regard to payment for the privilege of reprinting articles from its publications was discussed. It was VOTED that normally the fee for reprinting an entire article should not be less than \$100, payable to the author.

The Secretary raised the question of having a monograph history of the Association prepared and distributed to members. Following a discussion, the Secretary was instructed to explore the possibility with Professor A. W. Coats. The Secretary was also instructed to investigate the possibility of developing an "oral history" program, consisting of recordings by prominent members of the economics profession.

A letter from Professor Jacques Drèze was read, asking the Executive Committee to consider the advisability of extending junior membership privileges to students registered in foreign universities. Following a discussion, it was VOTED to adopt this recommendation.

It was VOTED to grant the request of Omicron Delta Epsilon, the honorary economics fraternity, to schedule a session at the 1964 annual meeting in Chicago and have the session listed in the joint program.

## REPORT OF THE SECRETARY FOR THE YEAR 1963

*Annual Meetings.* The final report of the 1962 meeting in Pittsburgh indicated a total registration of 4,528, of which 2,651 were members of the American Economic Association. The net income from the meeting was \$14,588.18. Of this amount, the American Economic Association received \$7,477.77, based on the number who registered as members.

The schedule for future annual meetings is: 1964, Chicago, Conrad Hilton Hotel; 1965, New York City, New York Hilton; 1966, San Francisco, San Francisco Hilton; 1967, Washington, D.C., Sheraton Park Hotel; 1968, Chicago (hotel to be confirmed); 1969, New York City, New York Hilton; and 1970, Detroit (hotel to be confirmed).

Mr. E. T. Baughman, Vice-President of the Federal Reserve Bank of Chicago, has agreed to serve as Chairman of the Committee on Local Arrangements for the 1964 meeting in Chicago.

*Membership.* Exhibit I below shows that the total number of members and subscribers was 16,279 as of November 30, 1963, a net increase of 635 for the year.

*Advertising and Announcements.* The number of advertising pages in the *American Economic Review* for 1963 was 155½ paid and 56 exchange pages compared with 128½ and 64½ pages in 1962. The "Vacancies and Applications" announcements totaled 30 pages for 1963 compared to 25 for 1962.

*Papers and Proceedings.* A comparison of the size and cost of the *Papers and Proceedings* for the period 1957-63 is made in Exhibit II below.

*The 1964 Handbook.* The preparation of the 1964 *Handbook* is completed and is scheduled for distribution to members during January, 1964.

*Permissions to Reprint and Translate.* The number of permissions to quote from, reprint, or translate articles from the *Review* and the *Papers and Proceedings* totaled 192 for 1963 compared to 220 for 1962. Under the policy adopted by the Executive Committee at the March, 1962, meeting, permission to reprint or translate is granted only after the consent of the author has been obtained by the editor or publisher.

*Use of Mailing List.* The list of institutions that used our mailing list in 1963 appears below. Requests for the use of the mailing list continue to come largely from publishers of books and periodicals and corporations wishing to send out reports or reprints of speeches. The sale of the mailing list continues to be an important source of income for the Association.

*Record Keeping.* The transfer of record keeping from the data processing company to the office of the Association was completed during 1963.

*Group Flights.* Plans were made for group flights to the Boston meeting from Chicago, Cincinnati, Denver, Pittsburgh, Los Angeles, and San Francisco. While there were sufficient reservations only for a group flight from Chicago to Boston, it is proposed to make similar arrangements for future meetings.

*Association Archives.* Considerable progress was made in establishing the Association archives during 1963. All material, including correspondence and committee reports, dealing with major policy questions has been retained. It is planned to arrange the files in chronological order to facilitate the work of those interested in using the archives.

*Asia Foundation Grant.* The Asia Foundation has given an additional sum of \$2,500 to enable the Association to continue assistance to Asian students.

*National Science Foundation Register of Scientific and Technical Personnel.* As authorized by the Executive Committee at its March, 1963, meeting, the American Economic Association made a contract with the National Science Foundation to participate in the Foundation's National Register of Scientific and Technical Personnel. Data for the register will be obtained by sending out a questionnaire to members of the economics profession in mid-March, 1964. The list will include all members of the A.E.A. plus others in the Southern Economic Association, Midwestern Economics Association, Western Economic Association, and the American Farm Economic Association. All returns will be coded in Evanston; then shipped to Raleigh, North Carolina, where the answers will be processed on the Foundation's IBM equipment. The information will, of course, always be at the Association's disposal. It is also understood that the Association will undertake to prepare a report based on the data.

*Clearinghouse Register.* Under a grant from the Ford Foundation to develop a roster of economists interested in overseas assignments, the Association sent out a pilot mailing on October 28 to members in the state of Michigan. Indicative of the interest of the members of the profession was a 25 percent return from this mailing of 425 questionnaires. On the basis of the pilot mailing, questionnaires were sent out to all members of the Association and others in the profession on December 15. It is expected that returns from the mailing will be classified and available to prospective employers by the end of February, 1964.

### *Standing Committees*

#### CENSUS ADVISORY COMMITTEE

Solomon Fabricant, *Chairman*  
(1963)

Harold Barger (1963)

Carl Kaysen (1963)

H. Gregg Lewis (1963)

Morris A. Adelman (1964)

Edward F. Denison (1964)

Werner Z. Hirsch (1964)

Edgar M. Hoover (1964)

Robert R. Nathan (1964)

Joe S. Bain (1965)

Bert G. Hickman (1965)

Guy H. Orcutt (1965)

Ralph W. Pfouts (1966)

Sherman J. Maisel (1966)

Gideon Rosenbluth (1966)

#### COMMITTEE ON ECONOMIC EDUCATION

Ben W. Lewis, *Chairman* (1963)

John R. Coleman (1963)

Edgar O. Edwards (1963)

Emanuel T. Weiler (1963)

Clark C. Bloom (1964)

Floyd A. Bond (1964)

Laurence Leamer (1964)

#### COMMITTEE ON HONORS AND AWARDS

Bernard F. Haley, *Chairman*  
(1966)

Gardner Ackley (1964)

Kenneth J. Arrow (1964)

Fritz Machlup (1964)

Earl J. Hamilton (1966)

Martin Bronfenbrenner (1966)

#### COMMITTEE ON RESEARCH AND PUBLICATIONS

Irving B. Kravis, *Chairman*  
(1965)

G. H. Hildebrand (1963)

Bert F. Hoselitz (1966)

Karl A. Fox (1964)

Franco Modigliani (1965)

Williard L. Thorp (1965)

Harold F. Williamson, *Ex Officio*



INSTITUTE OF INTERNATIONAL EDUCATION, ADVISORY AND POLICY BOARD REPRESENTATIVES	Benjamin Higgins (1966) Anthony Tang (1966)
Theodore Morgan, <i>Chairman</i> (1963)	J.C.E.E., SPECIAL PUBLICATIONS ADVISORY COMMITTEE
Carter Goodrich (1963)	G. L. Bach
Lorie Tarshis (1963)	K. E. Boulding
Andrew Kamarck (1965)	Ben W. Lewis
John Sheahan (1965)	James Reese
	Leonard Silk

*Committees Appointed During the Year*

COMMITTEE ON ELECTIONS	John A. Adler
Richard B. Heflebower, <i>Chairman</i>	Bernard F. Haley
John H. Wills	Richard Ruggles
Harold F. Williamson, <i>Ex Officio</i>	Robert A. Solow
COMMITTEE ON THE FUTURE ROLE OF THE A.E.A. IN ECONOMIC EDUCATION	FINANCE COMMITTEE
Ben W. Lewis, <i>Chairman</i>	C. Wells Farnham, <i>Chairman</i>
G. L. Bach	Corliss D. Anderson
R. A. Gordon	James Washington Bell
Abba P. Lerner	Harold F. Williamson
W. Allen Wallis	NOMINATING COMMITTEE
Harold F. Williamson	Paul A. Samuelson, <i>Chairman</i>
COMMITTEE ON THE INTERNATIONAL ECONOMIC ASSOCIATION	Robert E. Baldwin
Howard S. Ellis, <i>Chairman</i>	Harry C. Eastman
	Edgar O. Edwards
	Frank W. Fetter
	Ilse Mintz

*Council and Other Representatives*

A.A.A.S.	NATIONAL ACADEMY OF SCIENCES— NATIONAL RESEARCH COUNCIL
Bert F. Hoselitz (1963)	Robert Dorfman
A.C.L.S.	Andreas G. Papandreou
Overton H. Taylor (1966)	N.B.E.R.
I.E.A.	Willard L. Thorp (1965)
Bernard F. Haley (1965)	NATIONAL SCIENCE FOUNDATION
Robert M. Solow (1968)	O. H. Brownlee
J.C.E.E., BOARD OF TRUSTEES	S.S.R.C.
Ben W. Lewis (1966)	G. H. Hildebrand (1963)
Frank De Vyver (1965)	Karl A. Fox (1964)
Henry Wallich (1965)	Franco Modigliani (1965)

*Representatives of the Association on Various Occasions*

CENTENNIAL CONVOCATION OF BOSTON COLLEGE
Jaroslav Vanek

EIGHTEENTH NATIONAL CONFERENCE OF THE NATIONAL COMMISSION ON  
TEACHER EDUCATION AND PROFESSIONAL STANDARDS

Arthur D. Lynn, Jr.

FIFTEENTH ANNIVERSARY CELEBRATION OF BETA GAMMA SIGMA

Hiram S. Davis

ASSOCIATION FOR HIGHER EDUCATION CONFERENCE

Ben W. Lewis

FIFTH CONFERENCE ON INTERNATIONAL EDUCATION

Goetz A. Briefs

NATIONAL ACADEMY OF SCIENCES CENTENNIAL CELEBRATION

Richard H. Holton

BATES COLLEGE, 100TH ANNIVERSARY CONVOCATION

James A. Storer

INAUGURATIONS

Thomas Marshall Hahn, Jr., Virginia Polytechnic Institute

Vernon O. Johns

Everett Needham Case, Colgate University

John Gambs

Thomas Hamilton, University of Hawaii

Shelley M. Mark

W. Allen Wallis, University of Rochester

Norman Kaplan

Foster J. Taylor, Louisiana Polytechnic Institute

John W. Chisholm

Keith Spalding, Franklin and Marshall College

Albert L. Bell

Robert Isaac White, Kent State University

Dallas M. Young

Dumont Francis Kenny, Queensborough Community College of the City University of New York

Harold Barger

Sanford Soverhill Atwood, Emory University

Sherman Dallas

*Use of the Mailing List*

Permission was granted to the following to use our mailing list:

ALDINE PUBLISHING COMPANY

AMERICAN ACADEMY OF POLITICAL AND SOCIAL SCIENCE

AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE

AMERICAN BANKERS ASSOCIATION

AFL-CIO

AMERICAN STATISTICAL ASSOCIATION

ATLANTIC MONTHLY

BRITISH PUBLICATIONS, INC.

BROOKINGS INSTITUTION

BULLETIN OF THE ATOMIC SCIENTISTS

CHALLENGE

COMMITTEE ON INTERNATIONAL EXCHANGE OF PERSONS  
DOW JONES & COMPANY  
FEDERAL RESERVE BANK OF BOSTON  
FORD FOUNDATION  
FORDHAM UNIVERSITY PRESS  
FORTUNE  
GENERAL ELECTRIC COMPANY  
HARPER AND ROW  
UNIVERSITY OF ILLINOIS, BUREAU OF ECONOMIC AND BUSINESS RESEARCH  
INSTITUTE OF LIFE INSURANCE  
RICHARD D. IRWIN, INC.  
JOHNS HOPKINS PRESS  
JOHNSON REPRINT CORPORATION  
JOURNAL OF ECONOMIC ABSTRACTS  
JOURNAL OF PURCHASING  
AUGUSTUS M. KELLEY  
MCGRAW-HILL PUBLISHING COMPANY  
NAMES IN THE NEWS  
NATIONAL ASSOCIATION OF MANUFACTURERS  
UNIVERSITY OF THE STATE OF NEW YORK  
PENNSYLVANIA CONFERENCE OF ECONOMISTS  
FREDERICK A. PRAEGER, INC.  
PRUDENTIAL INSURANCE COMPANY OF AMERICA  
RANDOM HOUSE, INC.  
REGIONAL SCIENCE ASSOCIATION  
SPARTAN BOOKS  
TRANS WORLD AIRLINES  
TWENTIETH CENTURY FUND  
UNITED STATES—JAPAN TRADE COUNCIL  
UNITED STATES DEPARTMENT OF STATE  
UNITED STATES TREASURY, COMPTROLLER OF THE CURRENCY  
WESLEYAN UNIVERSITY PRESS  
WESTERN ECONOMIC ASSOCIATION

Respectfully submitted,  
HAROLD F. WILLIAMSON, *Secretary*

EXHIBIT I  
MEMBERS AND SUBSCRIBERS

	Totals 11/30/62*	Gain or Loss	Totals 11/30/63
Class of membership:			
Annual.....	9,557	595	10,152
Junior.....	1,298	33	1,331
Family.....	135	12	147
Complimentary.....	93	4	97†
Life.....	182	46	228
Honorary.....	20	2	18
Total members.....	11,285	688	11,973
Subscribers.....	4,359	53	4,306
Totals.....	15,644	635	16,279

\* These figures should not be considered definitive, since reports received from the data processing agency were incomplete.

† Includes 23 who do not receive publications.

EXHIBIT II  
PUBLICATION COSTS

PAPERS AND PROCEEDINGS				HANDBOOKS		
Year*	Number of pages	Number of copies	Cost	Number of pages	Number of copies	Cost
1957	754	12,400	\$16,253	548	10,100	\$15,815
1958	677	12,700	15,471	32	9,300	1,434
1959	689	14,000	16,780			
1960	745	14,800	18,914			
1961	675	15,800	19,759			
1962	615	16,000	18,277			
1963	753	17,700	23,440			

\* This is the year of publication and pertains to the meeting of the preceding year.

REPORT OF THE TREASURER FOR THE YEAR  
ENDING NOVEMBER 30, 1963

The comparative results of the operation of the Association for 1962 and 1963 are shown in Table 1. Higher subscription rates and membership dues, plus greater receipts from advertising, republication, and the sale of mailing lists resulted in an increase in income of just under \$41,000 during 1963 over 1962. Administrative and operating expenses of the Association for 1963 were only slightly above 1962; but largely because of an increase in the number of issues printed and the larger size of the *Papers and Proceedings*, the total publishing expense increased approximately \$8,500. The major item accounting for the deficit in net operating income of \$109 for 1963 was the expense of publishing the 1964 *Handbook*.

The comparative financial status of the Association for 1962 and 1963 is shown in Table 2. An advance of the first installment of funds needed for the preparation of Volume VI of the *Index of Economic Journals* and a contribution to the publication of the volume on Walras were largely responsible for the \$7,645 decline in unappropriated surplus during 1963.

Tables 3 and 4 give summaries of the investment portfolios and returns on investment for the years 1945-63. An alphabetical list of the Association's security holdings for 1963 is shown in Table 5.

The prospects for a more satisfactory net operating income for 1964 are encouraging, despite the possibility of somewhat higher costs of printing and administrative expenses. Further additions to membership and new subscribers should provide a moderate increase in income. But with the bulk of the publishing costs of the 1964 *Handbook* already paid and the final elimination of charges for outside accounting services, total income should exceed total expenses by a considerable margin.

Respectfully submitted,  
HAROLD F. WILLIAMSON, *Treasurer*

TABLE 1

COMPARATIVE RESULTS OF INCOME AND EXPENSES, 1962 AND 1963

	11/30/62	11/30/63
<i>Income</i>		
Membership dues.....	\$ 63,052	\$ 84,839
Subscriptions.....	26,823	35,260
Sales.....	4,092	3,850
Advertising.....	20,920	23,653
Republications income.....	7,692	12,574
Sale of mailing list.....	2,830	5,165
Sundry income.....	1,662	945
Dues and publications income.....	\$127,071	\$166,286
Interest.....	\$ 4,838	\$ 6,088
Dividends.....	4,489	5,041
Less custodian fees.....	372*	332*
Profit on sales of securities (net).....	2,587	2,349
Investment income (less fees).....	\$ 11,533	\$ 13,146
Total income.....	\$138,604	\$179,432
<i>Expenses</i>		
Office salaries.....	\$ 31,239	\$ 31,325
Accounting service (IBM agency).....	11,600	6,570
Other administrative expenses.....	10,707	13,274
Annual meeting.....	7,714*	5,626*
Executive Committee.....	1,286	2,645
Other committee expenses.....	1,249	876
Administrative and operating expenses.....	\$ 48,367	\$ 49,064
Review printing.....	\$ 43,780	\$ 49,941
Papers and Proceedings printing.....	18,277	23,440
Handbook printing.....	—	36,529
Editorial office (Review):		
Contributors.....	2,926	1,963
Editorial and clerical salaries.....	20,078	17,053
Other expenses (net).....	369	1,551
Publications.....	\$ 85,430	\$130,477
Total expenses.....	\$133,797	\$179,541
Net operating income or loss.....	\$ 4,807	\$ 109*

\* Denotes red.

TABLE 2  
COMPARATIVE FINANCIAL CONDITION, 1962 AND 1963

	11/30/62	11/30/63
<i>Assets</i>		
Cash on deposit and on hand.....	\$ 51,191	\$ 94,958
Receivables (net).....	12,714	13,919
Prepaid expenses and inventories.....	785	841
Furniture and fixtures (net).....	1,762	6,417
Investments at cost:		
Bonds.....	125,367	90,227
Stocks.....	116,699	119,524
Total assets.....	\$308,518	\$325,886
<i>Liabilities and Funds</i>		
Accounts payable.....	\$ 11,856	\$ 15,767
Deferred income.....	14,323	18,004
Committee on Research and Publications.....	353	32
Committee on Economic Education.....	108	—
Committee on <i>Papers and Proceedings</i> .....	178	—
Outside grants:		
Clearinghouse for economists (Ford).....	—	33,285
National register (N.S.F.).....	—	957*
Register of economists (Ford).....	1,836	74
Economic abstracts (Ford).....	8,197	12,600
Social studies survey (Ford).....	12,681	—
Foreign economic research translations (Ford).....	25,000	20,694
Foreign economic surveys (Ford).....	39,085	36,689
Survey articles (Rockefeller).....	3,250	1,625
Travel grant (Carnegie).....	1,579	1,578
Asia Foundation grant.....	1,188	728
Sundry.....	644	172
Life memberships.....	18,600	23,600
Total liabilities and funds.....	\$138,878	\$163,891
<i>Surplus</i>		
Balance at beginning of period.....	\$174,778	\$169,640
Expenditures in excess of grant, <i>Index of Economic Journals</i> .....	9,945*	5,536*
Contribution for publication of Jaffe's Life of L. Walras.....	—	2,500*
Cancellation of appropriation for Committee on Professional Standards and Ethics.....	—	500
Net income or loss for period.....	4,807	109*
Unappropriated surplus.....	\$169,640	\$161,995
Total footings.....	\$308,518	\$325,886

\* Denotes red.

TABLE 3  
INVESTMENT PORTFOLIO

YEAR	AT PAR	COST			MARKET
	Bonds	Bonds	Stocks	Total	Stocks and Bonds
1945	\$ 40,000	\$ 36,705	\$ 44,955	\$ 81,661	\$103,574
1948	35,000	33,108	48,624	81,732	84,841
1950	35,000	33,108	51,978	85,087	104,177
1951	43,000	43,340	49,764	93,104	117,316
1952	42,000	42,312	58,934	101,246	130,836
1953	68,000	68,308	46,458	114,766	134,562
1954	61,000	61,518	38,082	99,600	132,280
1955	75,000	75,370	59,394	134,764	166,772
1956	75,000	75,370	60,237	135,607	168,337
1957	75,000	75,370	55,084	130,454	151,638
1958	75,000	75,370	67,741	143,111	175,609
1959	75,000	75,386	67,652	143,038	191,506
1959*	175,000	175,616	67,652	243,268	291,506
1960*	160,000	160,508	94,910	255,418	299,768
1961*	170,000	169,794	109,071	278,865	356,131
1962*	125,000	125,367	116,699	242,066	293,039
1963*	90,000	90,367	119,524	209,891	284,160

\* Includes bonds held in temporary operating fund.

TABLE 4  
RETURN ON INVESTMENTS

Year	Bonds	Stocks	Total	Rate of Return on Cost
1945	\$1,479	\$2,488	\$3,968	4.71%
1948	1,194	2,944	4,139	5.06
1950	1,117	3,860	4,977	5.85
1951	1,026	4,607	5,633	6.05
1952	1,117	3,681	4,799	4.75
1953	1,435	3,587	5,022	4.36
1954	1,621	2,961	4,582	4.58
1955	1,750	3,002	4,752	3.53
1956	1,770	3,336	5,106	3.76
1957	1,770	3,397	5,167	3.90
1958	1,770	3,182	4,952	3.46
1959*	2,518	3,231	5,749	3.90
1959†	3,894	3,231	7,125	2.90
1960†	6,693	3,772	10,465	4.09
1961	5,460	4,143	9,603	3.44
1962	4,838	4,489	9,327	3.85
1963‡	3,320	5,041	8,361	3.98

\* Does not include income from bonds held in temporary operating fund.

† Includes income from bonds held in temporary operating fund.

‡ Does not include interest on savings account.



TABLE 5  
LIST OF SECURITIES HELD BY THE ASSOCIATION  
Stocks

Number of Shares of Stock	Issues	Cost	Approximate Market Value 11/30/63
100	Abbott Laboratories	\$ 6,133	\$ 11,500
150	American Potash and Chemical Corp.....	6,284	4,650
300	Central & South West Corp.....	2,101	12,600
150	Chain Belt Co.....	6,621	6,150
200	Continental Illinois National Bank & Trust Co...	6,619	8,400
100	Deere & Co.....	4,240	6,800
100	Farbenfabriken-Bayer A. G.....	6,822	6,300
200	Gulf Oil Corp.....	1,393	9,000
300	Houston Lighting & Power Co.....	1,240	12,600
100	Inland Container Corp.....	4,944	4,900
30	International Business Machines Corp.....	9,086	14,550
100	International Nickel Co. of Canada.....	3,911	6,400
100	McIntyre Porcupine Mines, Ltd.....	4,818	4,100
100	Marsh & McLennan, Inc.....	3,685	3,500
150	Montgomery Ward & Co.....	5,013	5,100
100	Motorola.....	6,746	8,200
200	Olin Mathieson Chemical Corp.....	8,731	8,600
220	Peoples Gas Light and Coke Co.....	3,562	11,000
100	Siemens and Halske.....	5,519	6,500
150	Socony Mobil Oil Co.....	3,882	9,750
100	Standard Oil Co. (Indiana).....	3,650	6,000
100	Swedish Ball Bearing.....	9,023	8,400
100	Wells Fargo Bank—San Francisco.....	3,521	8,200
150	Zenith Radio Corp.....	1,977	12,000
		\$119,521	\$195,200

## Bonds and Savings Account

Par Amount		Cost	Approximate Market Value 11/30/63
\$50,000	Savings Deposit 4%.....	\$ 50,000	\$ 50,000
5,000	U.S. Treas. Notes 3½% "D-1964" 5/15/64.....	5,000	5,000
8,000	U.S. Treas. Notes 1½% EO-1964 10/1/64.....	8,000	8,000
50,000	U.S. Treas. Bonds 3½% due 5/15/66.....	50,000	50,000
20,000	U.S. Treas. Bonds 3½% due 11/15/71.....	20,000	19,800
7,000	U.S. Treas. Bonds 2½% due 12/15/72-67.....	7,275	6,160
	Cash and Bonds.....	\$140,275	\$138,960
	Stocks.....	119,521	195,200
	Total.....	\$259,796	\$334,160

## REPORT OF THE FINANCE COMMITTEE

December 23, 1963

*Executive Committee  
American Economic Association  
Evanston, Illinois*

GENTLEMEN:

The accompanying tables show the list of securities held by the Association at the end of the fiscal year, November 30, 1963, and the changes made during that year. Table 1 gives a classified list of the stocks and the bonds according to maturity and records the cost and approximate market values on November 30, 1963. The securities account of the Association does not include "temporary operating funds" which are deposited in a savings account and approximated \$24,000 at the end of the last fiscal year.

The total market value of the securities account as of November 30, 1963, was \$334,160. This compares with \$278,039 at the end of the previous year. However, it should be noted that during the last fiscal year \$30,000 was transferred from temporary operating funds to the investment account and that only about \$26,000 of the increase in value was due to appreciation.

It has been the practice of the Association to set aside a portion of its resources consisting of grants made for specific purposes subject to early withdrawal to finance projects underway. These resources have been invested in U. S. Treasury securities or held in savings accounts. The transfer of \$30,000 to the securities account during the past fiscal year will leave approximately \$24,000 in the "temporary operating fund," which under present conditions seems adequate.

You will note from Table 2 that few changes were made in the securities account during the year. The Finance Committee remains optimistic as to the trend of general business and corporate profits and believes that the almost certain tax reduction may stimulate rising prices. Therefore, the Committee believes that a fairly aggressive investment position is desirable in spite of the admittedly high level of the stock market which would make it quite vulnerable to any disappointing developments.

Respectfully submitted,  
C. WELLS FARNHAM, *Chairman*  
CORLISS D. ANDERSON  
JAMES WASHINGTON BELL  
HAROLD F. WILLIAMSON

TABLE 1

INVENTORY AND APPRAISAL OF SECURITIES AND CASH AS OF NOVEMBER 30, 1963

	Par or Shares	Cost	Market Value
<b>FIXED INCOME SECURITIES</b>			
<i>Cash Equivalent</i>			
Savings Deposit 4%.....	\$50,000	\$ 50,000	\$ 50,000
U.S. Treas. 3½ 5/15/64.....	5,000	5,000	5,000
U.S. Treas. 1½ 10/1/64.....	8,000	8,000	8,000
U.S. Treas. 3½ 5/15/66.....	50,000	50,000	50,000
		\$113,000	\$113,000
<i>Bonds</i>			
U.S. Treas. 3½ 11/15/71.....	\$20,000	20,091	19,800
U.S. Treas. 2½ 12/15/72-67.....	7,000	7,276	6,160
		\$ 27,367	\$ 25,960
<b>TOTAL FIXED INCOME SECURITIES</b>		<b>\$140,367</b>	<b>\$138,960</b>
<b>COMMON STOCKS</b>			
<i>Utilities</i>			
Central & South West.....	300	2,101	12,600
Houston Lighting & Power.....	300	1,240	12,600
Peoples Gas.....	220	3,562	11,000
			\$ 36,200
<i>Financial</i>			
Continental Illinois National Bank.....	200	6,619	8,400
Wells Fargo Bank.....	100	3,522	8,200
Marsh & McLennan.....	100	3,687	3,500
			\$ 20,100
<i>Merchandising</i>			
Montgomery Ward.....	150	5,013	5,100
<i>Paper and Textiles</i>			
Inland Container.....	100	4,944	4,900
<i>Machinery and Construction</i>			
Chain Belt.....	150	6,621	6,150
Deere & Company.....	100	4,240	6,800
			\$ 12,950
<i>Mining and Metals</i>			
International Nickel.....	100	3,911	6,400
McIntyre Porcupine Mines.....	100	4,818	4,100
			\$ 10,500
<i>Oil and Gas</i>			
Gulf Oil.....	200	1,393	9,000
Socony Mobil.....	150	3,882	9,750
Standard Oil of Indiana.....	100	3,650	6,000
			\$ 24,750
<i>Chemicals and Drugs</i>			
Abbott Laboratories.....	100	6,133	11,500
American Potash.....	150	6,284	4,650
Olin Mathieson.....	200	8,731	8,600
			\$ 24,750
<i>Electrical Products</i>			
Motorola.....	100	6,746	8,200
Zenith Radio.....	150	1,977	12,000
			\$ 20,200

TABLE 1 (continued)

	Par or Shares	Cost	Market Value
<i>Office Equipment</i>			
International Business Machines.....	30	9,086	14,550
<i>Foreign</i>			
Bayer, A. G.....	100	6,822	6,300
Siemens & Halske.....	100	5,519	6,500
Swedish Ball Bearing.....	100	9,023	8,400
			\$ 21,200
TOTAL COMMON STOCKS.....		\$119,524	\$195,200
TOTAL SECURITIES.....		\$259,891	\$334,160
TEMPORARY OPERATING FUND			
Savings Deposit.....	\$ 23,895		

TABLE 2  
SUMMARY OF SECURITIES PURCHASED AND SOLD  
YEAR ENDED NOVEMBER 30, 1963

		Cost	Proceeds	Profit or Loss*
<i>Sales</i>				
	Shares			
Stocks—				
Castle and Cooke.....	104	\$ 3,672	\$ 3,118	\$ 554*
Monsanto Chemical Co.....	158.10	5,474	7,772	2,299
Wells Fargo Bank.....	21	740	1,268	528
		\$ 9,885	\$12,158	\$2,273
	Par Value			
Bonds—				
U.S. Treas. Notes 3½% 11/15/63....	\$35,000	\$35,000	\$35,077	\$ 77
Exchanged for others bonds—				
U.S. Treas. Bonds 2½% 8/15/63...	\$20,000	\$20,091		
<i>Purchases</i>				
Stocks—				
Marsh and McLennan.....	100	\$ 3,687		
Swedish Ball Bearing.....	100	9,023		
		\$12,710		
Bonds—				
Received in exchange for other bonds—				
U.S. Treas. Bonds 3½% 11/15/71..	\$20,000	\$20,091		
Less—cash received on exchange..		140		
		\$19,951		

\* Denotes red.

## REPORT OF THE AUDITOR

February 18, 1964

*Executive Committee  
American Economic Association  
Evanston, Illinois*

DEAR SIRs:

In accordance with instructions, we have examined the accounts and related records of the American Economic Association for the year ended November 30, 1963, and now submit our report thereon together with the following exhibits:

Statement of Financial Position—November 30, 1963	Exhibit 1
Statement of Income and Expense for year Ended November 30, 1963	Exhibit 2
Special Funds—Year Ended November 30, 1963	Exhibit 3

### *Results from Operations*

Net loss for the year ended November 30, 1963, was \$7,645 compared with net income of \$279 for the year ended November 30, 1962, as shown in the following summary:

Particulars	Year Ended November 30 1963	1962	Increase Decrease*
<b>Income:</b>			
Dues .....	\$ 84,839	\$ 63,052	\$ 21,787
Interest and dividends (net) .....	10,797	8,955	1,842
Profit on sale of securities (net) .....	2,349	2,577	228*
Miscellaneous income .....	945	1,662	717*
Total income .....	<u>\$ 98,930</u>	<u>\$ 76,246</u>	<u>\$ 22,684</u>
<b>Expense:</b>			
Publication expense .....	\$130,477	\$ 85,430	\$ 45,047
Less—Publication income .....	75,337	59,527	15,810
Net publication expense .....	\$ 55,140	\$ 25,903	\$ 29,237
Administrative and other operating expenses .....	43,899	45,536	1,637*
Appropriations (net) .....	7,536	4,528	3,008
Total expense .....	<u>\$106,575</u>	<u>\$ 75,967</u>	<u>\$ 30,608</u>
Net income or loss* .....	<u>\$ 7,645*</u>	<u>\$ 279</u>	<u>\$ 7,924*</u>

\*Denotes red.

Interest on bonds owned was accounted for in accordance with stated rates; dividends received on stocks were compared with amounts reported in published records as dividends paid.

Net publication expense, as shown in the following summary, amounted to \$55,140 for the current year compared with \$25,903 for the preceding year:

Particulars	Year Ended November 30	
	1963	1962
<b>Expenses:</b>		
Printing of—		
<i>Review</i> .....	\$ 49,941	\$ 43,780
<i>Directory and Handbook</i> .....	36,529	—
<i>Proceedings</i> .....	23,440	18,277
Editor's honorarium .....	8,000	8,000
Payments to contributors .....	1,963	2,926
Editorial clerical salaries .....	9,053	12,078
Editorial supplies and expense .....	1,551	369
Total expenses .....	<u>\$130,477</u>	<u>\$ 85,430</u>
<b>Less—Income:</b>		
Subscriptions, other than members .....	\$ 35,260	\$ 26,823
Sales of copies .....	3,850	4,092
Advertising .....	23,653	20,920
Republications .....	12,574	7,692
Total income .....	<u>\$ 75,337</u>	<u>\$ 59,527</u>
Net publication expense .....	<u>\$ 55,140</u>	<u>\$ 25,903</u>

Charges for the cost of printing the December, 1963, issue of the *Review* and reprints had not been made by the publishers at the time of our examination. The publishers estimate the cost of the *Review* printings and reprints at \$13,120; this amount is included in the expenses for the year.

### Financial Position

Financial position of the Association at November 30, 1963, and 1962 is set forth in the following summary:

	November 30		Increase Decrease*
	1963	1962	
<b>Assets</b>			
Cash on deposit and on hand .....	\$ 94,957	\$ 51,191	\$ 43,766
Receivables (net) .....	13,919	9,914	4,005
Advances to Foundation Project Chairmen .....	—	3,800	3,800*
Prepaid expenses .....	841	785	56
Equipment, furniture and fixtures (net) .....	6,417	1,762	4,655
Investments at cost—			
Bonds .....	90,227	125,367	35,140*
Stocks .....	119,525	116,699	2,826
	<u>\$325,886</u>	<u>\$309,518</u>	<u>\$ 16,368</u>
<b>Liabilities, Funds and Surplus</b>			
Accounts payable .....	\$ 15,767	\$ 11,856	\$ 3,911
Deferred income .....	18,004	14,323	3,681
Funds .....	106,520	95,099	11,421
Life memberships .....	23,600	18,600	5,000
Unappropriated surplus—			
Balance at beginning of year .....	169,640	169,361	279
Net income or loss* for year .....	7,645*	279	7,924*
	<u>\$325,886</u>	<u>\$309,518</u>	<u>\$ 16,368</u>

\*Denotes red.

Cash on deposit was satisfactorily reconciled with balances confirmed directly to us by the depositories.

The receivables of the Association were not confirmed by correspondence with debtors. Based upon the Association's past experience the reserve for doubtful accounts appears to be adequate to cover normal losses.

Changes in the investment account were verified by the examination of broker's invoices and other supporting data. Securities held at November 30, 1963, were confirmed directly to us by the State Bank and Trust Company of Evanston, Illinois, custodian for the Association.

#### *Funds*

The receipts and expenditures during the year for the various funds and grants are detailed in Exhibit 3. Total funds available for specific projects increased to \$106,519.58 at November 30, 1963.

We express our appreciation for the courtesies and cooperation extended our representatives during the course of the examination.

Very truly yours,  
DAVID HIMMELBLAU & Co.  
*Certified Public Accountants*



## EXHIBIT 1

AMERICAN ECONOMIC ASSOCIATION  
STATEMENT OF FINANCIAL POSITION—NOVEMBER 30, 1963

<i>Assets</i>		<i>Liabilities, Funds and Surplus</i>	
<b>CURRENT ASSETS:</b>		<b>CURRENT LIABILITIES:</b>	
Cash on deposit and on hand—		Accounts payable.....	\$ 15,767.35
State Bank and Trust Company, Evanston—		DEFERRED INCOME:	
Commercial account.....	\$ 21,001.37	Prepaid subscriptions.....	\$ 13,432.00
Savings account.....	73,896.12	Prepaid dues.....	4,572.00
Petty cash.....	60.00		18,004.00
	\$ 94,957.49		
<b>Receivables—</b>		<b>SPECIAL FUNDS (Exhibit 3).....</b>	<b>106,519.58</b>
Review advertising.....	\$ 6,361.00		
Accrued interest and dividends.....	910.94		
Publication sales, etc.....	6,547.83		
Membership dues.....	290.00		
	\$ 14,109.77		
Less—Reserve for doubtful accounts.....	190.44		
	13,919.33		
<b>Inventory of postage.....</b>	<b>352.53</b>		
<b>Unexpired insurance and prepaid expense.....</b>	<b>488.47</b>		
	\$109,717.82	<b>LIFE MEMBERSHIPS AND SURPLUS:</b>	
Total current assets.....		Life memberships.....	\$ 23,600.00
<b>INVESTMENTS (at cost):</b>		Unappropriated surplus—	
Bonds (market value—\$88,672.50).....	\$ 90,227.10	Balance November 30,	\$169,640.02
Stocks (market value—\$195,670.00)....	119,524.36	1962.....	
		Net loss for year ended	
<b>EQUIPMENT, FURNITURE AND FIXTURES</b>		November 30, 1963	
(less accumulated depreciation).....	6,417.13	(Exhibit 2).....	7,644.54
			185,595.48
	\$325,886.41		\$325,886.41

## AUDITORS' OPINION

*Executive Committee  
American Economic Association:*

In our opinion, the accompanying financial statements present fairly the financial position of the American Economic Association at November 30, 1963, and the results of its operations for the year ended that date, in conformity with generally accepted accounting principles applied on a basis consistent with that of the preceding year.

Our examination was made in accordance with generally accepted auditing standards and included such tests of the accounting records and other auditing procedures as we considered necessary in the circumstances.

Chicago, Illinois  
February 7, 1964

DAVID HIMMELBLAU & Co.  
Certified Public Accountants

## EXHIBIT 2

AMERICAN ECONOMIC ASSOCIATION  
STATEMENT OF INCOME AND EXPENSE  
FOR THE YEAR ENDED NOVEMBER 30, 1963

	Particulars	Amount		
INCOME:				
Dues—				
	Regular, junior and family members.....	\$76,089.56		
	Subscribing and contributing members.....	8,749.50	\$84,839.06	
Investments—				
	Interest on bonds and savings account.....	\$ 6,088.27		
	Dividends.....	5,041.07		
		\$11,129.34		
	Less custodian fees.....	332.52	10,796.82	
	Gain on sale of securities.....		2,349.23	
	Miscellaneous income.....		944.86	
	Total income.....		\$98,929.97	
EXPENSE:				
Administrative and other expense—				
	Secretary's salary.....	\$ 7,775.04		
	Office salaries.....	23,550.14		
	Addressing service income less expense.....	5,165.36*		
	Accounting service (IBM service).....	6,570.00		
	Postage.....	2,129.72		
	Stationery and printing.....	2,511.97		
	Insurance.....	237.44		
	Executive committee expense.....	2,645.19		
	Other committee expense.....	875.62		
	Annual meeting income less expense.....	5,626.52*		
	Annuity payments.....	2,448.44		
	Social security taxes.....	1,657.02		
	Provision for depreciation.....	831.97		
	Telephone and telegraph.....	552.62		
	International Economic Association.....	400.00		
	Office supplies.....	494.07		
	Miscellaneous expense (net).....	2,011.22	\$43,898.58	
Publication expense—				
Printing of—				
	Review.....	\$ 49,941.24		
	Directory and Handbook.....	36,528.36		
	Proceedings.....	23,440.17		
	Editorial honorarium.....	8,000.00		
	Payments to contributors.....	1,963.00		
	Editorial clerical salaries.....	9,052.66		
	Editorial supplies and expense.....	1,551.24		
		\$130,476.67		
Less—Publication income:				
	Subscriptions other than members.....	\$35,259.62		
	Sales of copies.....	3,850.35		
	Advertising.....	23,653.00		
	Republications.....	11,573.77		
	Reprints.....	1,000.00	75,336.74	55,139.93
				99,038.51
	Excess of expenses over income (forward).....			\$ 108.54

\* Denotes red.

## EXHIBIT 2 (continued)

ADD:

Expenditures in excess of grant for Cumulative <i>Index of Economic Journals</i> for the year ended November 30, 1963.....	\$ 5,536.00(A)	
Contribution for the publication of Jaffe's Life of Louis Walras	2,500.00	
	<u>\$ 8,036.00</u>	
Less cancellation of appropriation for Committee on Professional Standards and Ethics.....	500.00	7,536.00
Net loss to surplus (Exhibit 1).....		<u>\$ 7,644.54</u>

(A) This represents 50% of the proposed budget of \$11,071.00 for publishing Volume VI of the *Index*.

## EXHIBIT 3

AMERICAN ECONOMIC ASSOCIATION  
STATEMENT OF SPECIAL FUNDS  
FOR YEAR ENDED NOVEMBER 30, 1963

<i>Fund</i>	Balance November 30, 1962	Received	Expended	Refunded to Grantor	Balances Charged Off	Balance November 30, 1963
• Carnegie Corporation of New York grant for travel expenses of delegates to international meetings.....	\$ 1,578.64	\$ —	\$ —	\$ —	\$ —	\$ 1,578.64
Rockefeller Foundation grant for survey articles on recent developments.....	3,250.00	1,625.00	3,250.00	—	—	1,625.00
The Ford Foundation grants—						
Preparation of a special register of economists.....	1,835.50	—	1,761.77	—	—	73.73
Translation of books and articles of major significance into the English language.....	25,000.00	—	4,305.46	—	—	20,694.54
Preparation and publication of articles surveying economic research in foreign countries.....	39,085.07	—	2,396.45	—	—	36,688.62
Survey study of economic text books in secondary schools.....	12,680.91	—	4,015.62	8,665.29	—	—
Fund for <i>Journal of Economic Abstracts</i> .....	9,197.45	20,000.00	16,597.55	—	—	12,599.90
Clearinghouse for Economists.....	—	35,000.00	1,715.07	—	181.49*	33,284.93
Committee on Economic Education.....	107.80	—	289.29	—	—	—
Committee on Papers and Proceedings.....	178.30	—	178.30	—	500.00	—
Committee on Professional Standards and Ethics.....	500.00	—	—	—	—	727.74
Asia Foundation.....	1,188.24	—	460.50	—	—	31.62
Committee on Publication and Research.....	353.17	—	321.55	—	—	957.14*
National Science Foundation (A).....	—	—	957.14	—	—	172.00
Sundry.....	144.00	100.00	72.00	—	—	—
Total funds (Exhibit 1).....	<u>\$95,999.08</u>	<u>\$56,725.00</u>	<u>\$36,320.70</u>	<u>\$8,665.29</u>	<u>\$318.51</u>	<u>\$106,519.58</u>

(A) Grant of \$18,600.00 for Economic Section of National Roster of Scientific and Technical Personnel. Expenditures are to be billed to Foundation on a monthly basis.

\* Denotes red.

# REPORT OF THE MANAGING EDITOR FOR THE YEAR ENDING DECEMBER 1963

The number of manuscripts received during 1963 was 329, which is substantially higher than last year's figure of 273. Table 1 gives comparative figures for the past five years.

TABLE 1  
MANUSCRIPTS RECEIVED 1959-63

	1963	1962	1961	1960	1959
Manuscripts received.....	329	273	305	276	279
Articles.....	142	157	179	158	180
Communications.....	187	116	126	118	99
Percentage of articles accepted.....	10	15	12	16	14

Table 2 provides the breakdown of the volume's contents between articles, review articles, communications, book reviews, etc. The allocation of space between the various sections of the *Review* has remained about the same in recent years, except that this year no review articles were published.

The 1963 volume was about the same size as that of 1962, and each of these was larger than the 1961 volume. This is primarily accounted for by the fact that each of the last two volumes included two survey articles instead of one (as in 1961). The survey articles are largely financed by the Rockefeller Foundation, so they are treated as net additions to the volume.

TABLE 2  
SUMMARY OF CONTENTS 1961-63

	1963		1962		1961	
	Number	Pages	Number	Pages	Number	Pages
Leading articles.....	19	508	18	442	20	441
Review articles.....			5	72	2	32
Communications:						
Original.....	11	66	7	65	10	66
Comments and replies...	16	57	16	84	15	48
Book reviews.....	209	372	197	386	210	377
Classified lists:						
New books.....		71		68		70
Periodicals.....		74		50		60
Dissertations.....		39		33		41
Notes.....		46		52		33
		1233*		1252*		1168

\*Plus some blank pages.

The two survey articles published in 1963, the sixth and seventh of the series of eight, were: "Regional Economics: A Survey," by John Meyer, and "Survey of Inflation Theory," by Martin Bronfenbrenner and Franklyn Holzman. The eighth article, on "The Theory of Bargaining," by Robert Bishop, will be published either in June or in September of next year. The Rockefeller grant, which was to have terminated in October, 1963, was extended for one year.

Table 3 shows the subject-matter distribution of all manuscripts submitted during 1963. More than 25 percent of all articles and original communications received were in the two fields of price theory and income theory. The "field" of economic history, development, and national economies was represented by 42 manuscripts (15 percent of the total), and money and banking by 30 manuscripts (11 percent). International trade and industrial organization were also popular fields.

TABLE 3  
SUBJECT-MATTER DISTRIBUTION OF MANUSCRIPTS SUBMITTED IN 1963

	Articles	Original Communi- cations	Comments; Replies	Totals
General economics.....	3	6	—	9
Price theory.....	13	25	6	44
Income theory.....	18	21	11	50
History of economic thought.....	3	1	—	4
Economic history; development; na- tional economies.....	28	14	5	47
Social accounting.....	1	3	—	4
Economic systems.....	4	2	2	8
Business fluctuations.....	2	1	2	5
Money and banking.....	15	15	3	33
Public finance.....	3	10	—	13
International economics.....	11	13	7	31
Business finance.....	1	1	1	3
Business organization.....	3	3	1	7
Industrial organization.....	15	4	11	30
Land economics; housing.....	2	4	1	7
Labor economics.....	10	5	—	15
Consumption; welfare; population.....	6	1	1	8
Unclassified.....	4	7	—	11
	142	136	51	329

Table 4 summarizes the subject-matter distribution of accepted articles, review articles, and communications for the past six years; the figures in parentheses give the distribution for 1963 only. The most interesting figures are those for leading articles plus original communications, since to some extent these figures indicate the areas in which most work of publishable quality and of broad interest to economists is being done. The five fields showing the highest concentration for the last six years are: price and allocation theory (28), income and employment theory (18), economic develop-

ment and national economies (17), international economics (15), and money and banking (10).

Table 5 presents the expenditures in 1963 for the four regular issues of the *Review* in comparison with the estimated budget and with the actual expenditures in 1962. The budget figures for 1963 did not include the estimated

TABLE 4  
SUBJECT-MATTER DISTRIBUTION OF ACCEPTED MANUSCRIPTS, 1958-63 AND 1963

	Articles	Review Articles	Original Communications	Comments; Replies	Totals
General economics.....	4	2	3	3	12
Price theory.....	19 (2)	3	9 (4)	5	36 (6)
Income theory.....	12 (1)	2	6 (2)	16 (5)	36 (8)
History of economic thought	4		1	2	7
Economic history; develop- ment; national economies.	13 (1)	3	4	15 (1)	35 (2)
Social accounting.....	3	1	1	1	6
Economic systems.....	3			2 (2)	5 (2)
Business fluctuations.....	7	1	2 (1)	3 (2)	13 (3)
Money and banking.....	6 (1)	3	4	10 (2)	23 (3)
Public finance.....	2	1	6 (2)	8	17 (2)
International economics....	12 (1)	2	3	6 (2)	23 (3)
Business finance.....	4 (1)			6 (1)	10 (2)
Business organization.....	5 (2)				5 (2)
Industrial organization....	7 (2)		2 (1)	10 (4)	19 (7)
Land economics; housing...	3	1	2		6
Labor economics.....	5 (1)		1	2	8 (1)
Consumption; welfare; population.....	6 (1)		1	3 (1)	10 (2)
Unclassified.....	1		5 (1)	1	7 (1)
	116 (13)	19 (0)	50 (11)	93 (20)	278 (44)

NOTE: The 1958-63 figure is followed in each case by the 1963 figure in parentheses.

TABLE 5  
ACTUAL AND BUDGETED EXPENDITURES

	Budget 1963	Actual 1963	Actual 1962
Printing and mailing .....	\$48,500	\$51,465.12	\$47,785.86*
Editor's salary.....	8,000	8,000.00	8,000.00
Assistant editor.....	3,000		
Acting editors†.....			1,500.00
Editorial assistance.....	9,650	9,702.06	10,385.05
Supplies (including postage).....	970	1,165.50	851.95
Contributors.....	2,500	2,308.00	3,375.00
Office equipment and decorating....		812.63	
	\$72,620	\$73,453.31	\$71,897.86
		Less: 3,250.00	Less: 3,250.00
		\$70,203.31	\$68,647.86

\*Corrected from 1962 Report.

† For 3 months.

expense of publishing two survey articles, but the actual expenditure figures for the year do include this expense. The survey-articles grant from the Rockefeller Foundation provided \$3,250 for the two survey articles. If this amount is subtracted from actual expenditures, the result is \$70,203.31, or \$2,416.69 less than the estimated budget. The amount of \$3,000 in the estimated budget for an Assistant Editor was not used during the year; on the other hand, \$812.63 was spent for office equipment and decorating, and this was not allowed for in the estimated budget.

Table 6 gives detailed information about printing cost by issues during 1963. The number of copies printed in 1963 averaged 18,050; in 1962, 16,125; in 1961, 15,650; in 1960, 15,000; in 1959, 14,125; and in 1958, 12,975. The successive increases are of course mainly related to the increases in membership and subscriptions.

TABLE 6  
COPIES PRINTED, SIZE, AND COST OF PRINTING IN 1963

	COPIES PRINTED	PAGES		COST		COST INCLUDING REPRINTS
		Net	Gross	Issue	Reprints	
March.....	17,700	282	352	\$11,790.75	\$118.38	\$11,909.13
June.....	18,000	309	360	12,433.70	106.35	12,540.05
September..	18,000	347	404	13,795.75	100.19	13,895.94
December..	18,500	295	365†	13,020.00†	100.00†	13,120.00†
		1,233*	1,481	\$51,040.20	\$424.92	\$51,465.12

\* Plus some blank pages.

† Estimated.

The estimated costs for the coming year are presented in Table 7, based on a volume of 1,420 pages, including advertising (or about 1,150 pages of text, exclusive of one survey article) and an average number of copies of 19,000. Anticipated cost of the survey article is not included, since this cost will be substantially covered by the Rockefeller Foundation grant. A salary increase of \$500 for Miss Ladd and \$2,000 for a book review editor are included in the amount for editorial and clerical assistance. The sum of \$3,000 for an Assistant Editor is not included this year.

TABLE 7  
RECOMMENDED BUDGET FOR 1964

Printing (including paper, postage, reprints) .....	\$51,000
Managing editor's salary .....	8,000
Editorial and clerical assistance .....	12,500
Supplies and postage .....	1,175
Contributors .....	1,600
Office equipment and decoration .....	170
	<hr/> \$74,445

During the year I have had the advice and assistance of the following foreign correspondents, who have been particularly helpful with regard to the selection of foreign books for listing and review:



Maurice Flamant (France)	P. J. Verdoorn (Netherlands)
Erich Schneider (Germany)	Bent Hansen (Sweden)
Oscar Soberon (Mexico)	

Two members of the Board of Editors complete their three-year terms of office at this time: Alfred E. Kahn and Joseph A. Pechman. The Association owes them a heavy debt of gratitude for the generous expenditure of time they have made in the interests of the *Review*, and I very much appreciate their constant willingness, even when other obligations pressed, to review manuscripts and offer editorial advice. I nominate as their successors on the Board, for three-year terms beginning in 1964: Arthur Goldberger and Richard Caves.

During the year I have frequently sought the aid of members of the profession in addition to the members of the Editorial Board—partly to relieve the latter of what would otherwise be an impossibly heavy burden and partly to obtain advice of specialists in particular areas not represented on the Board. The following have assisted in this way:

I. Adelman	R. Dorfman	L. Klein	D. Orr
A. Alchian	E. Edwards	L. Krause	A. Papandreou
S. Alexander	R. Eisner	M. Kreinin	R. Powell
A. Ando	R. Evans	R. Kuenne	R. Quandt
G. C. Archibald	D. Fand	E. Kuh	G. Ranis
K. Arrow	J. C. H. Fei	M. Kurz	M. Reder
R. Attiyeh	W. Fellner	S. Kuznets	L. S. Ritter
H. Averch	R. Ferber	S. Lebergott	G. Rosenbluth
G. L. Bach	M. Frankel	M. L. Lee	H. Rosovsky
J. Bain	V. Fuchs	F. de Leeuw	J. Schmookler
F. Bator	L. E. Gallaway	R. Lester	T. Scitovsky
W. Baumol	G. Garvy	J. Letiche	R. Selden
P. W. Bell	F. Gehrels	H. S. Levine	M. Shubik
S. T. Beza	A. Goldberger	H. Leibenstein	H. Simon
G. Borts	R. Goode	R. Lipsey	V. Smith
G. Break	Z. Griliches	F. Machlup	W. Smith
K. Brunner	G. Hadley	R. McKinnon	E. Solomon
J. Buchanan	D. Hamberg	A. Manne	R. Solow
P. Cagan	L. Hansen	J. Margolis	J. Stein
P. Cartwright	R. Heflebower	G. Meier	P. Steiner
R. Caves	L. Herman	J. Mellor	G. Stigler
L. V. Chandler	B. Hickman	M. Michaely	L. Tarshis
H. Chenery	J. Hirschleifer	M. Miller	J. Tobin
C. Christ	F. Holzman	E. S. Mills	H. Uzawa
P. David	L. J. Johnson	H. Minsky	H. Wagner
E. Denison	D. Jorgenson	R. Mundell	H. Wallich
E. Despres	A. Kahan	R. Musgrave	J. Wiseman
D. Dillard	J. Kareken	M. Nerlove	L. Yeager
E. Domar	C. Kindleberger	D. North	

Respectfully submitted,  
JOHN G. GURLEY, *Managing Editor*

## REPORT OF THE COMMITTEE ON RESEARCH AND PUBLICATIONS

The sixth volume of the *Index of Economic Journals*, covering the period through 1962, is in process of preparation.

In the "Readings Series," volumes on welfare economics, edited by Kenneth Arrow and Tibor Scitovsky, and on business cycles, edited by R. A. Gordon and Lawrence R. Klein, are well under way. The Committee is investigating the possibility of volumes in price theory, international trade, and agricultural economics. The first two subjects have already had volumes devoted to them but expert opinion is being sought to determine whether the time is ripe for a second collection of articles in each of these fields.

The series of translations of the work of foreign economists, supported by a grant from the Ford Foundation, has been set in motion thanks to Professor Bert Hoselitz. Commitments have been made for the translation of three works: (1) a collection of 16 Chinese papers written from 800 to 1500 dealing with monetary theory, (2) Dahmen's *Svensk Industriell Forstagarverksamhet*, and (3) Chayanov's *Die Lehre von der buerlichen Wirtschaft*. The Committee is considering several other titles but it has unanimously agreed that the most important foreign economist whose work has not been adequately translated into English is Pareto. Accordingly, a search is being made for a suitable translator and the Committee has decided that it is willing to devote the major portion of its unexpended Ford grant for a scholarly translation of the *Manuel* with suitable annotations involving comparisons with the earlier Italian version and the *Cours*.

IRVING B. KRAVIS, *Chairman*

## REPORT ON THE JOURNAL OF ECONOMIC ABSTRACTS

One complete volume of four numbers of the *Journal of Economic Abstracts* has now been published and Volume II, Number 1, will appear in January, 1964. Volume I abstracted from 33 journals and contained abstracts of 649 articles written by 739 authors.

Subscriptions, flooding in at first, have now tapered off as would be expected and come at a rate of about fifty or sixty per week. Subscribers, numbered 6,075, as of December 15, 1963, of which about two-thirds are located in the United States. Receipts from subscriptions, including prepaid renewals, net of foreign exchange and bank service charges, totaled \$12,886.16 as of December 15, 1963. Total costs were \$27,909.04. Disbursed from the Ford grant have been \$20,737.78, leaving a balance of \$19,262.22. The *Journal* has a \$5,714.90 balance on hand.

ARTHUR SMITHIES, *Editor*

## REPORT OF THE COMMITTEE ON THE FUTURE ROLE OF THE A.E.A. IN ECONOMIC EDUCATION

The Committee recommends Executive Committee approval of the following:

1. The adoption and publication of a policy statement indicating the concern of the Association with the problem of economic education.

2. The reorganization of the Association's Committee on Economic Education into a smaller, more action-oriented committee, to be concerned with economic education at the precollege and "adult" levels as well as the collegiate level.

3. The development of closer working relations with the Joint Council on Economic Education. Beyond the Association's current affiliation and nomination of three members for the Joint Council's Board of Trustees, a portion of the membership of the reorganized Committee on Economic Education should be established as a special advisory committee to the Joint Council, to advise the Council on its operations in the field of education, research, and the preparation of better teaching materials for classroom use. It would be understood that this does not in any way make the American Economic Association responsible for the activities of the Joint Council, but would provide close, active liaison and greater participation by professional economists in the work of the Joint Council.

4. The establishment of channels for more effective assistance by professional economists throughout the country to local school systems interested in improving economics in their classes, probably utilizing the services of local Councils of the Joint Council on Economic Education.

5. The appointment of a part-time assistant to work with the chairman of the new Committee on Economic Education, particularly in developing more effective relationships with the Joint Council and in carrying out the kinds of action programs approved by the committee.

BEN W. LEWIS, *Chairman*

## COMMITTEE ON THE INTERNATIONAL ECONOMIC ASSOCIATION

The two principal activities of the International Economic Association since September, 1962, have been (1) the conference on activity analysis held in Cambridge, England, at the end of June, 1963, and (2) the conference on the economics of education held at Menthon St. Bernard, France, at the end of August, 1963, in connection with the meeting of the Executive Committee of the Association.

The conference on activity analysis was the second conference primarily for younger economists. There were twenty-one participants, including seven from the United States. Robert Dorfman was a member of the program committee for the conference.

The conference on the economics of education was attended by forty-two economists from sixteen different countries: Belgium, Canada, Chile, Czechoslovakia, France, Germany, Great Britain, Greece, Italy, Nigeria, Pakistan, Poland, Sweden, Turkey, United States, U.S.S.R. There were eight participants from the United States. The conference extended over a period of nine days. Mary Jean Bowman was a member of the program committee.

Plans are being made for three conferences to be held in 1964: (1) a conference primarily for younger economists to be held in Israel in March on price formation in different economies; (2) the annual September conference at Palermo, Italy, on general factors governing the distribution of national income; and (3) a regional conference on problems of the economic development of socialist countries in Eastern Europe, organized with the financial help of UNESCO, to be held in December at Varna, Bulgaria.

At the August meeting of the Association's Executive Committee, on which Bernard F. Haley currently represents the A.E.A., the Secretary General made it clear that the Association is now confronted with serious financial problems; and it was agreed that the constituent national associations should be informed of this situation and their executive committees asked to support the measures proposed for meeting these problems.

The major problem arises from the fact that the grant of \$35,000 a year from the Ford Foundation terminates with the year 1964. This five-year grant of \$175,000 made early in 1960 was the second grant to the Association by the Foundation, the first one (of \$115,000) having been made in 1954. It is now necessary for the officers of the International Economic Association again to approach the Ford Foundation with an application for a further grant, since without such a grant the conferences and other activities of the Association would have to be severely curtailed. The importance of this source of income is indicated by the fact that, in the draft budget for 1964, it provides about 45 percent of anticipated income.

When the second grant was made by the Ford Foundation in 1960, officers of the Foundation expressed the hope that the I.E.A. would now find other sources of income so that the Foundation would not continue to be so predominantly the main support of the Association. Since 1960, accordingly, the

Executive Committee of the I.E.A. has sought contributions from central banks and has in addition obtained substantial financial assistance from various sources for individual conferences. Contributions from central banks have amounted to more than \$4,000 in 1963, and the Association expects to raise about \$10,000 from this source in 1964. Various grants and services made available to the Association without charge amount to the equivalent of \$10,000 to \$20,000 a year. In addition, at its last meeting the Executive Committee increased the dues of the constituent member associations by 50 percent to a total amount of about \$20,000. It is to be hoped that the amount of these supplementary sources of income will put the Association in a strong position in approaching the Ford Foundation.

Another financial problem for the Association concerns the regular annual contribution of UNESCO, amounting to \$9,000. There is a serious possibility that this grant may be discontinued as a result of a general movement on the part of a number of national delegations to UNESCO, particularly those outside Europe, who take the view that international associations of more than ten years standing have had time to obtain financing from other sources. They advocate limiting UNESCO subventions to special contracts, mostly concentrated on the needs of underdeveloped countries. This point of view found considerable support at UNESCO's last general conference, and the question is on the agenda for action at the 1964 conference.

It is important to the I.E.A. that the annual subvention from UNESCO should be continued, if only because this additional source of support strengthens the case that can be made to the Ford Foundation.

Given these two major financial problems, the Executive Committee of the I.E.A. requests supporting action from the member associations. Specifically, the Executive Committee of the American Economic Association is asked:

1. To support the application to the Ford Foundation for a renewal of its grant of \$35,000 a year.

2. To authorize the Association's Committee on the International Economic Association to approach the Federal Reserve Bank of New York with a view to obtaining a contribution for the support of the I.E.A.

3. To support a request, directed to appropriate officers in the U.S. Department of State, that the U.S. delegation to the general conference of UNESCO favor continuance of the annual grant of \$9,000 to the International Economic Association.

The Committee on the International Economic Association recommends that the Executive Committee give its support in the ways requested.

In the accompanying Annex a summary of the activities of the International Economic Association is provided for the information of the Executive Committee.

HOWARD S. ELLIS, *Chairman*  
JOHN A. ADLER  
BERNARD F. HALEY  
RICHARD RUGGLES  
ROBERT A. SOLOW

## ANNEX

The International Economic Association was brought formally into existence in 1950. It is an association of national economics associations, and there are currently forty member associations. The purpose of the Association is to provide opportunities, through conferences and publications, for the interchange of ideas among economists of different countries. An annual September conference has been held each year beginning in 1950 on some topic in economic theory or policy of current importance and interest. About seventy-five different U.S. economists have participated in these conferences, not counting about a hundred others who attended the 1962 Congress of the Association in Vienna. Nearly all of these conferences have resulted, or will result, in published volumes of the papers and discussions; some distinguished examples are the following: *Monopoly and Competition and Their Regulation*, *The Business Cycle in the Post-War World*, *The Theory of Wage Determination*, *The Theory of Capital Inflation*. Nine volumes of this series have now appeared, and four others are in process.

In addition to the regular annual conferences, there have been two full congresses of the Association to which all individual members of the constituent associations have been invited: one at Rome in 1956 on "Stability and Progress in the World Economy," and the other at Vienna in 1962 on "Economic Development." (These were substituted for the regular September conferences in these two years.)

Aided by the two grants from the Ford Foundation, the Association was able to expand its range of activities, beginning in 1957, by holding a series of regional conferences on problems of economic development in different parts of the world: Latin America (meeting at Rio de Janeiro in 1957); East Asia (Gamagori, 1960); and Africa South of the Sahara (Addis Ababa, 1961). Beginning in 1962, an annual conference for younger economists was instituted, and two of these have now been held. In addition there have been two special conferences. Altogether thirty-three different U.S. economists have participated in the regional conferences, those for younger economists, and the special conferences.

Another important activity of the Association has been a series of "refresher courses" designed primarily for younger teachers in underdeveloped countries. Four of these have been held: in Poona; Singapore; Murvee (Pakistan); Manila; and Addis Ababa. Six U.S. economists have aided in this work.

In addition to the published volumes growing out of the annual September conferences, the regional conferences have also resulted, or will result, in published proceedings, as will also the conferences for younger economists. These various publications have been self-financing. The Association has also performed a most useful function by sponsoring *International Economic Papers*, translations into English of important foreign-language articles of historical and contemporary interest, eleven volumes of which have now appeared.

From the beginning of the International Economic Association in 1950, the American Economic Association has been a strong supporter of the organization. The first two presidents of the new Association were Americans. Over a hundred different U.S. economists have been active participants in the conferences held by the Association. Continued support by the A.E.A. is most important to the growth and strength of the international association.

# REPORT OF ADVISORY AND POLICY BOARD REPRESENTATIVES INSTITUTE OF INTERNATIONAL EDUCATION THE 1963 ECONOMICS INSTITUTE

Members of the Policy and Advisory Board during 1963 were: Theodore Morgan, Chairman (1963), Carter Goodrich (1963), Lorie Tarshis (1963), Andrew Kamarck (1965), John Sheahan (1965), Benjamin Higgins (1965), and Anthony Tang (1966).

In 1963, 83 applications to attend the Institute were submitted to the Director. Of these 69 were offered admission, and 55 attended—the largest number so far.

The students came from 28 countries, the most from Japan (11), Colombia (6), Argentina (4), France (3), and Brazil, Egypt, Greece, India, Korea, Nepal, and Spain (2 each). In September they went off to 27 different universities and colleges, the largest numbers to Vanderbilt (6), Chicago (5), Colorado, Columbia, Harvard, and Stanford (4 each), Williams and Yale (3 each), and Cornell, Pennsylvania, and Wisconsin (2 each).

The Institute program consisted of (1) a survey of macro- and micro-economic theory, offered at two levels of difficulty, and a series of lectures on the U. S. economy; (2) instruction in the English language, at three levels, concentrated in the earlier weeks; (3) supplementary work in mathematics and statistics; and (4) recreation and orientation.

In June of 1963, the Ford Foundation made a grant of \$250,000, which it is estimated should cover the net costs of operating the Institutes through 1968 and possibly 1970.

SUMMARY OF COSTS OF THE ECONOMIC INSTITUTES—1958-63

YEAR	NUMBER OF STUDENTS	EXPENSES			
		Adminis- trative Expense*	Other Expense†	Total Expense	Total Expense per Student
1958.....	36	\$12,395	\$ 35,451	\$ 47,846	\$1,329
1959.....	49	8,453	45,745	54,198	1,106
1960.....	46	6,326	47,781	54,107	1,176
1961.....	53	4,509	52,778	57,287	1,081
1962.....	50	3,000	57,704	60,704	1,214
1963.....	55	3,370	63,246	66,616	1,211
Totals.....	289	\$38,053	\$302,705	\$340,758	
Average expense per student....					\$1,179

\* Administrative expense consists of off-campus charges, paid directly by the I.I.E.: administrative overhead, pre- and post-Institute orientation and travel, Policy Board costs, and other direct charges. Starting up expenses made the first year's outlay especially large. There was no meeting of the Board in 1962, which lowered costs for that year.

† On-campus charges.



YEAR	NUMBER OF STUDENTS	REVENUES			COST TO FORD GRANT	
		Department of State*	Other Sponsors*	Total Revenue	Total	Cost per Student
1958.....	36	\$ 4,982 (5)	\$ 4,263 (11)	\$ 9,244	\$ 38,602	\$1,072
1959.....	49	7,750 (20)	7,661 (10)	15,411	38,787	792
1960.....	46	6,975 (18)	9,819 (15)	16,794	37,313	811
1961.....	53	7,363 (9)	11,252 (15)	18,614	38,673	730
1962.....	50	9,123 (16)	6,200 (7)	15,323	45,381	908
1963.....	55	7,000 (14)	19,264 (21)	26,264	40,352	733
Totals.....	289	\$43,193	\$58,459	\$101,650	\$239,108	
Averages per student....		527 (82)	740 (79)			827

\* Figures in parentheses are the numbers of students for whom their sponsors made payments.

THEODORE MORGAN

## REPORT OF THE CENSUS ADVISORY COMMITTEE

The Census Advisory Committee of the American Economic Association met with the Director and Staff of the Census Bureau in January, 1963. The Secretary and Assistant Secretary of Commerce also participated during part of the meeting. Principal topics discussed were: recent developments in statistical programs; long-range program and selection of priority items for early budget presentation; the new survey of manufacturers' sales, orders, and inventories; new uses of the annual survey of manufactures; expanded data on capacity and investment; statistical requirements related to the Trade Expansion Act (1962) and the Tariff Classification Act (1962); and Census Bureau plans for research and development.

In addition, the Technical Subcommittee on Business Cycle Developments, a subcommittee of the Census Advisory Committee, met with the Director and Staff in May. Among the topics discussed were: recent improvements in business cycle developments; smoothing techniques; research on amplitude-adjusted composite indexes; research on financial series; possible companion report on long-term economic growth; suggestions for further improvement in business cycle developments; and the treatment of inverted series.

Members of the Census Advisory Committee are:

Morris A. Adelman  
Harold Barger  
Edward F. Denison  
Solomon Fabricant, *Chairman*  
Bert G. Hickman  
Werner Z. Hirsch  
Edgar M. Hoover  
Carl Kaysen

H. Gregg Lewis  
John Lintner  
Sherman J. Maisel  
Robert R. Nathan  
Guy H. Orcutt  
Ralph W. Pfouts  
Gideon Rosenbluth

Members of the Technical Subcommittee on Business Cycle Developments are:

Donald J. Daly  
Gottfried Haberler  
Bert G. Hickman, *Chairman*  
Lawrence R. Klein  
John P. Lewis

Geoffrey H. Moore  
Frank E. Morris  
Arthur M. Okun  
Beryl W. Sprinkel  
Lorman C. Trueblood  
SOLOMON FABRICANT, *Chairman*

## REPORT OF REPRESENTATIVE TO THE NATIONAL BUREAU OF ECONOMIC RESEARCH

The program of the National Bureau of Economic Research in 1963 included research in five areas: Economic Growth; National Income, Consumption, and Capital Formation; Business Cycles; Financial Institutions and Processes; and International Economic Relations. Nine reports were published in 1963.

*Directors, Officers and Staff.* At the 1963 Annual Meeting Francis M. Boddy was elected a member and Director by Appointment of the University of Minnesota. Theodore O. Yntema was elected a member and Director by Appointment of the Committee for Economic Development. Nathaniel Goldfinger was elected a member and Director by Appointment of the American Federation of Labor and Congress of Industrial Organizations. Jacob Viner was elected Director Emeritus.

Officers elected at the 1963 Annual Meeting were Albert J. Hettinger, Jr., Chairman; Arthur F. Burns, President; Frank W. Fetter, Vice President; Donald B. Woodward, Treasurer; Solomon Fabricant, Director of Research; Geoffrey H. Moore and Hal B. Lary, Associate Directors of Research; and William J. Carson, Executive Director and Secretary.

*Finances.* The National Bureau's total research and operating expenditures in 1963 were \$1,236,923.

Several new grants were received in 1963. The National Science Foundation made one grant for an investigation of the performance of the U.S. in world trade. The office of Manpower, Automation and Training, U.S. Department of Labor, made a grant for a study of labor force trends and projections in the United States; the Ford Foundation made a grant for a study of productivity in service industries; International Business Machines Corporation made a grant for a study of application of electronic computers in the study of business cycles; and four finance companies made a joint grant to supplement an earlier grant for the study of consumer finance.

WILLARD L. THORP

# PUBLICATIONS

## OF THE

### AMERICAN ECONOMIC ASSOCIATION

1964

#### FIRST SERIES

*Numbers starred are sold only with the sets; the supply of those double starred is exhausted. For information apply to the Secretary.*

##### Volume I, 1886

	<i>Price in paper</i>
1. Report of Organization of the American Economic Association. Pp. 46.	\$ .50
2-3. **Relation of the Modern Municipality to the Gas Supply. By E. J. James. Pp. 66.	.75
4. Co-operation in a Western City. By Albert Shaw. Pp. 106.	.75
5. **Co-operation in New England. By E. W. Bemis. Pp. 136.	.75
6. **Relation of the State to Industrial Action. By H. C. Adams. Pp. 85.	.75

##### Volume II, 1887

1. Three Phases of Co-operation in the West. By Amos G. Warner. Pp. 119.	.75
2. **Historical Sketch of the Finance of Pennsylvania. By T. K. Worthington. Pp. 106.	.75
3. The Railway Question. By Edmund J. James. Pp. 68.	.75
4. **Early History of the English Woolen Industry. By W. J. Ashley. Pp. 85.	.75
5. **Mediaeval Guilds of England. By E. R. A. Seligman. Pp. 113.	.75
6. Relation of Modern Municipalities to Quasi-public Works. By H. C. Adams and others. Pp. 87.	.75

##### Volume III, 1888

1. Statistics in College, by C. D. Wright; Sociology and Political Economy, by F. H. Giddings; The Legal-Tender Decisions, by E. J. James. Pp. 80.	.75
2. **Capital and Its Earnings. By John B. Clark. Pp. 69.	.75
3. The Manual Laboring Class, by F. A. Walker; Mine Labor in the Hocking Val- ley, by E. W. Bemis; Report of the Second Annual Meeting. Pp. 86.	.75
4-5. **Statistics and Economics. By Richmond Mayo-Smith. Pp. 127.	1.00
6. The Stability of Prices. By Simon N. Patten. Pp. 64.	.75

##### Volume IV, 1889

1. **Contributions to the Wages Question: The Theory of Wages, by Stuart Wood; Possibility of a Scientific Law of Wages, by J. B. Clark. Pp. 69.	.75
2. **Socialism in England. By Sidney Webb. Pp. 73.	.75
3. Road Legislation for the American State. By J. W. Jenks. Pp. 83.	.75
4. Third Annual Meeting: Report of the Proceedings. Pp. 123.	.75
5. **Malthus and Ricardo, by S. N. Patten; The Study of Statistics, by D. R. Dewey; Analysis in Political Economy, by W. W. Folwell. Pp. 69.	.75
6. An Honest Dollar. By E. Benjamin Andrews. Pp. 50.	.50

## Volume V, 1890

- |  |        |
|--|--------|
| 1. The Industrial Transition in Japan. By Yeijiro Ono. Pp. 122.                      | \$1.00 |
| 2. Two Essays on Child-Labor. By W. F. Willoughby and Clare de Graffenried. Pp. 150. | .75    |
| 3-4. Papers on the Canal Question. By E. J. James and L. M. Haupt. Pp. 85.           | 1.00   |
| 5. History of the New York Property Tax. By J. C. Schwab. Pp. 108.                   | 1.00   |
| 6. Educational Value of Political Economy. By S. N. Patten. Pp. 36.                  | .75    |

## Volume VI, 1891

- |   |      |
|---|------|
| 1-2. Fourth Annual Meeting: Reports, Papers, Discussions.               | 1.00 |
| 3. Government Forestry. Papers by Pinchot, Bowers, and Fernow. Pp. 102. | .75  |
| 4-5. Municipal Ownership of Gas in the U.S. By E. W. Bemis. Pp. 185.    | 1.00 |
| 6. State Railroad Commissions. By F. C. Clark. Pp. 110.                 | .75  |

## Volume VII, 1892

- |  |      |
|--|------|
| 1. **The Silver Situation in the United States. By F. W. Taussig. Pp. 118.           | .75  |
| 2-3. **Shifting and Incidence of Taxation. By E. R. A. Seligman. Pp. 424. (Revised.) | 2.00 |
| 4-5. **Sinking Funds. By Edward A. Ross. Pp. 106.                                    | 1.00 |
| 6. The Reciprocity Treaty with Canada of 1854. By F. E. Haynes. Pp. 70.              | .75  |

## Volume VIII, 1893

- |  |      |
|--|------|
| 1. Fifth Annual Meeting: Report of the Proceedings. Pp. 130.                 | .75  |
| 2-3. Housing of the Poor in American Cities. By M. T. Reynolds. Pp. 132.     | 1.00 |
| 4-5. Public Assistance of the Poor in France. By E. G. Balch. Pp. 180.       | 1.00 |
| 6. **First Stages of the Tariff Policy of the U.S. By William Hill. Pp. 162. | 1.00 |

## Volume IX, 1894

- |   |      |
|---|------|
| Sixth Annual Meeting: Handbook and Report. Pp. 73.  | .50  |
| 1-2. **Progressive Taxation in Theory and Practice. By Edwin R. A. Seligman. Pp. 222. (See 1908, No. 4.)            | 1.00 |
| 3. **The Theory of Transportation. By C. H. Cooley. Pp. 148.  | .75  |
| 4. **Sir William Petty. By Wilson Lloyd Bevan. Pp. 102.   | .75  |
| 5-6. **Papers on Labor Problems, By J. B. Clark, C. D. Wright, D. R. Dewey, A. T. Hadley, and J. G. Brooks. Pp. 94. | .75  |

## Volume X, 1895

- |  |      |
|--|------|
| Seventh Annual Meeting: Handbook and Report. Pp. 183.                                  | .50  |
| 1-3. **The Canadian Banking System, 1817-1890. By R. M. Breckenridge. Pp. 478.         | 1.50 |
| 4. Poor Laws of Massachusetts and New York. By John Cummings. Pp. 136.                 | .75  |
| 5-6. **Letters of Ricardo to McCulloch, 1816-1823. Edited by J. H. Hollander. Pp. 204. | 1.75 |

## Volume XI, 1896

- |   |      |
|---|------|
| 1-3. **Race Traits and Tendencies of the American Negro. By F. L. Hoffman. Pp. 330. | 1.25 |
| 4. **Appreciation and Interest. By Irving Fisher. Pp. 110.                          | .75  |
| *General Index to Volume I-XI (1886-1896).  | .25  |

## ECONOMIC STUDIES

(\$2.50 per volume)

## Volume I, 1896

- |  |     |
|--|-----|
| Eighth Annual Meeting: Handbook and Report. Pp. 78.  | .50 |
| 1. **The Theory of Economic Progress, by J. B. Clark; The Relation of Changes in the Volume of the Currency to Prosperity, by F. A. Walker. Pp. 46.                    | .50 |
| 2. The Adjustment of Wages to Efficiency. Three Papers: Gain Sharing, by H. R. Towne; The Premium Plan, by F. A. Halsey; A Piece-Rate System, by F. W. Taylor. Pp. 83. | .50 |
| 3. **The Populist Movement. By Frank L. McVey. Pp. 81.   | .50 |
| 4. The Present Monetary Situation. By W. Lexis; translated by John Cummings. Pp. 72.   | .50 |
| 5-6. The Street Railway Problem in Cleveland. By W. R. Hopkins. Pp. 94.  | .75 |

## Volume II, 1897

- Ninth Annual Meeting: Handbook and Report. Pp. 162. \$ .50
1. \*\*Economics and Jurisprudence. By Henry C. Adams. Pp. 48. .50
  2. The Saloon Question in Chicago. By John E. George. Pp. 62. .50
  3. The General Property Tax in California. By C. C. Plehn. Pp. 88. .50
  4. Area and Population of the United States at the Eleventh Census. By W. F. Willcox. Pp. 60. .50
  5. \*\*A Discussion Concerning the Currencies of the British Plantations in America, etc. By William Douglass. Edited by C. J. Bullock. Pp. 228. .50
  6. Density and Distribution of Population in the United States at the Eleventh Census. By W. F. Willcox. Pp. 79. .50

## Volume III, 1898

- Tenth Annual Meeting: Handbook and Report. Pp. 136. .50
1. Government by Injunction. By William H. Dunbar. Pp. 44. .50
  2. Economic Aspects of Railroad Receiverships. By H. H. Swain. Pp. 118. .50
  3. The Ohio Tax Inquisitor Law. By T. N. Carver. Pp. 50. .50
  4. The American Federation of Labor. By Morton A. Aldrich. Pp. 54. .50
  5. Housing of the Working People in Yonkers. By E. L. Bogart. Pp. 82. .50
  6. The State Purchase of Railways in Switzerland. By Horace Michellie; translated by John Cummings. Pp. 72. .50

## Volume IV, 1899

- Eleventh Annual Meeting: Handbook and Report. Pp. 126. .50
1. I. Economics and Politics. By A. T. Hadley. II. Report on Currency Reform. III. Report on the Twelfth Census. Pp. 70. .50
  2. Personal Competition. By Charles H. Cooley. Pp. 104. .50
  3. Economics as a School Study. By F. R. Clow. Pp. 72. .50
  - 4-5. The English Income Tax. By J. A. Hill. Pp. 162. 1.00
  6. \*\*Effects of Recent Changes in Monetary Standards upon the Distribution of Wealth. By F. S. Kinder. Pp. 91. .50

## NEW SERIES

1. \*\*The Cotton Industry. By M. B. Hammond. Pp. 382. 1.50
2. Scope and Method of the Twelfth Census. Critical discussion by over twenty statistical experts. Pp. 625. 2.00

## THIRD SERIES

*Note—During 1896-1899 the Association issued its publications in two series, viz., the bimonthly Economic Studies, and the "New Series" of larger monographs printed at irregular intervals. In 1900 it reverted to the policy of issuing its monographs, now called the "Third Series" of the publications, at quarterly intervals.*

*Price per volume, \$4.00.*

## Volume I, 1900

1. Twelfth Annual Meeting: Papers on Economic Theory and Political Morality; Trusts; Railroad Problems; Public Finance; Consumers' League; Twelfth Census. Pp. 186. 1.00
2. \*\*The End of Villeinage in England. By T. W. Page. Pp. 99. 1.00
3. Essays in Colonial Finance. By Members of the Association. Pp. 303. 1.50
4. \*\*Currency and Banking in the Province of Massachusetts Bay. By A. McF. Davis. Part I: Currency. Pp. 464 + 91 photogravure plates. 1.75

## Volume II, 1901

1. \*\*Thirteenth Annual Meeting; Papers on Competition; Commercial Education; Economic Theory; Taxation of Quasi-public Corporations; Municipal Accounts. Pp. 300. 1.25
2. Currency and Banking. By A. McF. Davis. Part II: Banking. Pp. 341 + 18 photogravure plates. 1.75
3. \*\*Theory of Value before Adam Smith. By Hannah R. Sewall. Pp. 132. 1.00
4. Administration of City Finances in the U.S. By F. R. Clow. Pp. 144. 1.00

## Volume III, 1902

1. Fourteenth Annual Meeting: Papers on International Trade; Industrial Policy; Public Finance; Protective Tariff; Negro Problem; Arbitration of Labor Disputes in Coal Mining Industry; Porto Rican Finance; Economic History. Pp. 400. \$1.50
2. \*\*The Negro in Africa and America. By Joseph A. Tillinghast. Pp. 240. 1.25
3. Taxation in New Hampshire. By M. H. Robinson. Pp. 232. 1.25
4. \*\*Rent in Modern Economic Theory. By Alvin S. Johnson. Pp. 136. .75

## Volume IV, 1903

1. Fifteenth Annual Meeting: Papers and Discussions on Economic and Social Progress; Trade Unions and the Open Shop; Railway Regulation; Theory of Wages; Theory of Rent; Oriental Currency Problem. Pp. 298. 1.25
2. Ethnic Factors in the Population of Boston. By F. A. Bushee. Pp. 171. 1.00
3. History of Contract Labor in the Hawaiian Islands. By Catharine Coman. Pp. 74. .75
4. \*\*The Income Tax in the Commonwealth of the United States. By Delos O. Kinsman. Pp. 134. 1.00

## Volume V, 1904

Sixteenth Annual Meeting. Papers and Proceedings published in two parts.

1. Part I—Papers and Discussions on Southern Agricultural and Industrial Problems; Social Aspects of Economic Law; Relations between Rent and Interest. Pp. 240. 1.00
2. Part II—Papers and Discussion on The Management of the Surplus Reserve; Theory of Loan Credit in Relation to Corporation Economics; State Taxation of Interstate Commerce; Trusts; Theory of Social Causation. Pp. 203. 1.00
3. Monopolistic Combinations in the German Coal Industry. By Francis Walker. Pp. 340. 1.25
4. \*\*The Influence of Farm Machinery on Production and labor. By Hadley Winfield Quaintance. Pp. 110. .75

## Volume VI, 1905

Seventeenth Annual Meeting. Papers and Proceedings published in two parts.

1. Part I—Papers and Discussions on the Doctrine of Free Trade; Theory of Prices; Theory of Money; Open Shop or Closed Shop. Pp. 226. 1.00
2. Part II—Papers and Discussions on Government Interference with Industrial Combinations; Regulation of Railway Rates; Taxation of Railways; Preferential Tariffs and Reciprocity; Inclosure Movement; Economic History of the United States. Pp. 270. 1.00
3. \*\*The History and Theory of Shipping Subsidies. By R. Meeker. Pp. 230. 1.00
4. Factory Legislation in the State of New York. By F. R. Fairchild. Pp. 218. 1.00

## Volume VII, 1906

1. Eighteenth Annual Meeting: Papers and Discussions on The Love of Wealth and the Public Service; Theory of Distribution; Governmental Regulation of Railway Rates; Municipal Ownership; Labor Disputes; The Economic Future of the Negro. Pp. 325. 1.00
2. Railroad Rate Control. By H. S. Smalley. Pp. 147. 1.00
3. On Collective Phenomena and the Scientific Value of Statistical Data. By E. G. F. Gryzanovskii. Pp. 48. .75
- Handbook of the Association, 1906. Pp. 48. .25
4. The Taxation of the Gross Receipts of Railways in Wisconsin. By G. E. Snider. Pp. 138. 1.00

## Volume VIII, 1907

1. Nineteenth Annual Meeting: Papers and Discussion on Modern Standards of Business Honor; Wages as Determined by Arbitration; Commercial Education; Money and Banking; Western Civilization and Birth Rate; Economic History; Government Regulation of Insurance; Trusts and Tariff; Child Labor. Pp. 268. 1.00
2. Historical Sketch of the Finances and Financial Policy of Massachusetts from 1780 to 1905. By C. J. Bullock. Pp. 144. 1.00
- Handbook of the Association, 1907. Pp. 50. .25
3. The Labor Legislation of Connecticut. By Alba M. Edwards. Pp. 322. 1.00
4. The Growth of Large Fortunes. By G. P. Watkins. Pp. 170. 1.00

## Volume IX, 1908

- Handbook of the Association, 1908. Pp. 49. \$ .25
1. Twentieth Annual Meeting: Papers and Discussions on Principles of Government Control of Business; Are Savings Income; Agricultural Economics; Money and Banking; Agreements in Political Economy; Labor Legislation; Relation of the Federal Treasury to the Money Market; Public Service Commissions. Pp. 311. 1.25
  2. Chicago Traction. By R. E. Heilman. Pp. 131. 1.00
  3. Factory Legislation of Rhode Island. By J. K. Towles. Pp. 119. 1.00
  4. \*\*Progressive Taxation in Theory and Practice. Revised Edition. By E. R. A. Seligman. Pp. 334. 1.25

## Volume X, 1909

1. Twenty-first Annual Meeting: Papers and Discussions on The Making of Economic Literature; Collective Bargaining; Round Table on Accounting; Labor Legislation; Employers' Liability; Canadian Industrial-Disputes Act; Modern Industry and Family Life; Agricultural Economics; Transportation; Revision of the Tariff; A Central Bank; The National Monetary Commission; Capitalization of Public Service Corporations in Massachusetts. Pp. 432. 1.50
2. Handbook of the Association, 1909. Pp. 59. .25
3. \*\*The Printers. By George E. Barnett. Pp. 379. 1.50
4. Life Insurance Reform in New York. By W. H. Price. Pp. 95. .75

## Volume XI, 1910

1. Twenty-second Annual Meeting: Papers and Discussions on History of the Association; Observation in Economics; Economic Dynamics; Theory of Wages; Country Life; Valuation of Public Service Corporations; Trusts; Taxation. Pp. 386. 1.50
2. \*\*Handbook of the Association, 1910. Pp. 79. .25
3. The Child Labor Policy of New Jersey. By A. S. Field. Pp. 229. 1.25
4. The American Silk Industry and the Tariff. By F. R. Mason. Pp. 178. 1.00

## THE ECONOMIC BULLETIN

Published quarterly in 1908, 1909, and 1910, and containing personal notes, news of the economic world, announcements of special lines of investigation, and a classified and annotated bibliography of the current books and articles on economic subjects.

- Volume I, 1908. Four numbers. \$2.00; each .60  
 Volume II, 1909. Four numbers. \$2.00; each .60  
 Volume III, 1910. Four numbers. \$2.00; each .60

## FOURTH SERIES

## Volume I, 1911

The American Economic Review, \*\*March, \*\*June, September, and December

## \*\*Twenty-third Annual Meeting:

Papers and Discussions on the Significance of a Comprehensive System of Education; Money and Prices; The Ricardo Centenary; Accounting; Canals and Railways; Population and Immigration; Labor Legislation; Taxation; A Definition of Socialism; Competition in the Fur Trade. Pp. 388.

\*\*Handbook of the Association, 1911.

## Volume II, 1912

The American Economic Review, March, June, September, and December Supplement.—Twenty-fourth Annual Meeting:

Papers and Proceedings. The Economic Utilization of History; Tariff Legislation; The Federal Budget; Rural Conditions; Selection of Population by Migration; The Price Concept; An International Commission on the Cost of Living; Industrial Efficiency. Pp. 146.

## Volume III, 1913

The American Economic Review, March, June \*\*September, and \*\*December Supplement.—Twenty-fifth Annual Meeting:



Papers and Proceedings. Population or Prosperity; The Rising Cost of Living—Standardizing the Dollar; Banking Reform; Theories of Distribution; Farm Management; Governmental Price Regulation. Pp. 155.

**\*\*Supplement.**—Handbook of the Association, 1913.

#### Volume IV, 1914

The American Economic Review, **\*\*March**, June, September, and December  
Supplement.—Twenty-sixth Annual Meeting:

Papers and Proceedings. The Increasing Governmental Control of Economic Life; The Control of Public Utilities; Railroad Rate Making; Syndicalism; Trust Decisions and Business. Pp. 211.

#### Volume V, 1915

The American Economic Review, March, June, September, and **\*\*December**  
Supplement.—Twenty-seventh Annual Meeting:

Papers and Proceedings. Economics and the Law; Regulation of the Stock Exchanges; Market Distribution; Statistical Work of the United States Government; Relation of Education to Industrial Efficiency; The Effect of Income and Inheritance Taxes on the Distribution of Wealth; Public Regulation of Wages. Pp. 323.

#### Volume VI, 1916

The American Economic Review, **\*\*March**, **\*\*June**, September, and **\*\*December**

**\*\*Supplement.**—Twenty-eighth Annual Meeting:

Papers and Proceedings. The Apportionment of Representatives; Effect of the War on Foreign Trade; Budget Making and the Increased Cost of Government; Economic Costs of War; Economic Theorizing and Scientific Progress; The Role of Money in Economic Theory; Price Maintenance; Investment of Foreign Capital. Pp. 248.

**\*\*Supplement.**—Handbook of the Association, 1916.

#### Volume VII, 1917

The American Economic Review, March, June, September, and December  
Supplement.—Twenty-ninth Annual Meeting:

Papers and Proceedings. The National Point of View in Economics; Landed Property; Two Dimensions of Economic Productivity; Some Social Surveys in Iowa; The Land Problem and Rural Welfare; The Federal Farm Loan Act; Statistics of the Concentration of Wealth; Gold Supply at the Close of the War; Loans and Taxes in War Finance; Problems of Population after the War; Some Phases of the Minimum Wage. Pp. 275.

Supplement.—Index to the Publications, 1886-1910.

#### Volume VIII, 1918

The American Economic Review, March, **\*\*June**, September, and **\*\*December**

**\*\*Supplement.**—Thirtieth Annual Meeting:

Papers and Proceedings. Economic Reconstruction; Federal Taxes upon Income and Excess Profits; Land Utilization and Colonization; Federal Valuation of Railroads; Co-ordination of Employment Bureaus; Control of the Acquisition of Wealth; Motives in Economic Life; Price-Fixing Problems of Governmental Efficiency; Economic Alliances and Tariff Adjustments. Pp. 317.

#### Volume IX, 1919

The American Economic Review, **\*\*March**, **\*\*June**, September, and December

**\*\*Supplement.**—Thirty-first Annual Meeting:

Papers and Proceedings. Economists in Public Service; Interest on Investment a Manufacturing Cost Factor; Control of Marketing Methods and Costs; War and the Supply of Capital; War and the Rate of Interest; Index of the Cost of Living; Securing the Initiative of the Workman; A Legal Dismissal Wage; After-War Gold Policies; Foreign Exchange; Stabilizing the Dollar; Tenancy of Landed Property; Price-Fixing; Economic Theory in an Era of Readjustment; Psychology and Economics; The Open Door and

Colonial Policy; Reports of Committees on Foreign Trade and Purchasing Power of Money. Pp. 368.

\*\*Supplement No. 2—Report of the Committee on War Finance. Pp. 142.

\*\*Supplement No. 3—Handbook of the Association, 1919.

#### Volume X, 1920

The American Economic Review, \*\* March, June, \*\* September, and \*\* December

\*\* Supplement.—Thirty-second Annual Meeting:

Papers and Proceedings. Excess Profits Taxation; Germany's Reparation Payments; International Supervision over Foreign Investments; Results of a Balanced Industrial System; Employee's Representation in Management of Industry; Prices and Reconstruction; Banking Policy and Prices; Large-Scale Marketing; Reports of Committees on Foreign Trade, Co-ordination in Taxation, Census Advisory Committee. Pp. 278.

Supplement No. 2—Taxation of Excess Profits in Great Britain. Pp. 244.

#### Volume XI, 1921

The American Economic Review, March, June, September, and December

Supplement.—Thirty-third Annual Meeting:

Papers and Proceedings. The Railroad Situation; Our Foreign Trade Balance; Problems of the Bituminous Coal Industry; Traditional Economic Theory; Non-Euclidean Economics; Federal Taxation of Profits and Income; Teaching of Elementary Economics. Pp. 194.

Supplement No. 2.—Karelsen Prize Essays, on What Can a Man Afford? Pp. 118.

#### Volume XII, 1922

The American Economic Review, March, June, September, and December

Supplement.—Thirty-fourth Annual Meeting:

Papers and Proceedings. Business Cycles; American Trade Unionism; The Railroads; Workmen's Compensation; Federation in Central America; Teaching of Elementary Economics; The Chain Store Grocer; Economics and Ethics. Pp. 194.

\*\*Supplement No. 2—Handbook of the Association, 1922.

#### Volume XIII, 1923

The American Economic Review, March, \*\* June, September, and December

Supplement.—Thirty-fifth Annual Meeting:

Papers and Proceedings. Trend of Prices; Trend of Rate of Interest and Investment; Outlook for Wages and Employment; Overhead Costs; Commercial Rent and Profits; Labor Turnover; Factors in Wage Determinations; Income of Farmers; Large-Scale Production and Merchandising; Marketing Farm Products; Bureaus of Business Research. Pp. 293.

#### Volume XIV, 1924

The American Economic Review, \*\* March, \*\* June, September, and December

Supplement.—Thirty-sixth Annual Meeting:

Papers and Proceedings. International Trade and Commercial Policy; Railroad Consolidation; Economic Theory; Transportation; American Foreign Trade; Marketing. Pp. 192.

\*\*Supplement No. 2—Handbook of the Association, 1924.

#### Volume XV, 1925

The American Economic Review, March, June, September, and December

Supplement.—Thirty-seventh Annual Meeting:

Papers and Proceedings. The Economics of Advertising; Problems of Economic Theory; Transportation; Marketing; Giant Power; The Teaching of Business and Economics; Business Administration; Monetary Stabilization; Foreign Service Training; Highway Economics; Psychological Problems of Industry. Pp. 165.

\*\* Supplement No. 2.—Babson Prize Essay, on Forecasting the Price of Hogs. Pp. 22.

#### Volume XVI, 1926

The American Economic Review, March, June, September, and December

Supplement.—Thirty-eighth Annual Meeting:

Papers and Proceedings. Movement of Real Wages; Teaching of Economics; Consuming Power of Labor and Business Fluctuations; Economic Problems Involved in the Payment of International Debts; Economics and Geography; Agriculture in Our National Policy; Tariff Making; Trade Associations; Theory of Wages; Reducing the Costs of Marketing; Topics in Economic History; Railway Problems; Land Economics; Federal Reserve Policies. Pp. 353.

**\*\*Supplement No. 2—Handbook of the Association, 1926.**

#### Volume XVII, 1927

The American Economic Review, March, **\*\*June**, September, and December

**\*\*Supplement.—Thirty-ninth Annual Meeting:**

Papers and Proceedings. Economics of Prohibition; Economic History; Use of the Quantitative Method in the Study of Economic Theory; Present-Day Corporation Problems; American Practices Analogous to Foreign Controls over Raw Materials; Marketing; Interest Theory and Price Movements; Problem of Effective Public Utility Regulation; Immigration Restriction—Economic Results and Prospects; Family Budgets; Motor Transportation in the United States. Pp. 218.

Supplement No. 2.—Report of the Dinner in Honor of Professor John Bates Clark. Pp. 18.

#### Volume XVIII, 1928

**\*\*Supplement No. 2.—Handbook of the Association, 1928,**

**Supplement.—Fortieth Annual Meeting:**

Papers and Proceedings. Land Economics; Marketing; Present Status and Future Prospects of Quantitative Economics; Post-War Fluctuations of Commodity Prices; Relationship between Departments of Economics and Collegiate Schools of Business; Economic History; Simplification of the Federal Income Tax; Economic Significance of the Increased Efficiency of American Industry; An Approach to the Law of Production and Its Relation to the Welfare of the Wage-Earner; Meaning of Valuation; Railroad Valuation with Special Reference to the O'Fallon Decision; Interest Rates As Factors in the Business Cycle; Should the Debt Settlements Be Revised; An Examination of the Reasons for Revision of the Debt Settlements. Pp. 305.

#### Volume XIX, 1929

The American Economic Review, March, June, September, and December.

**Supplement.—Forty-first Annual Meeting:**

Papers and Proceedings. Market Shifts, Price Movements, and Employment; Some Observations on Unemployment Insurance; Marketing; Land Economics; Law and Economics; Price Stabilization; London and the Trade Cycle; Federal Reserve Policy and Brokers' Loans; Central Planning of Production in Soviet Russia; International Differences in the Labor Movement; Tariff Making in the United States; Economic History; Locality Distribution of Industries; Regulation of Electric Light and Power Utilities; An Inductive Study of Publicly Owned and Operated vs. Privately Owned but Regulated Public Utilities; Regulation of the Common Carrier; Commercial Motor Vehicle and the Public. Pp. 284.

#### Volume XX, 1930

The American Economic Review, March, June, September, and December

**Supplement.—Forty-second Annual Meeting:**

Papers and Proceedings. Economic History; Public Works Plan and Unemployment; Theory of Economic Dynamics as Related to Industrial Instability; Chief Economic Problems of Mexico; Reparations Settlement and the International Flow of Capital; Federal Reserve Board—Its Problems and Policy; Economic and Social Consequences of Mechanization in Agriculture and Industry. Pp. 214.

#### Volume XXI, 1931

The American Economic Review, March, June, September, and December

**Supplement.—Forty-third Annual Meeting:**

Papers and Proceedings. Decline of Laissez Faire; Small Loan Business; Social and Economic Aspects of Chain Stores; Russian Economic Situation; Trustification and Economic Theory; Persistence of the Merger Movement; Program of the Federal Farm Board; Social Implications of Restriction of Agricultural Output; Land Economics and Real Estate; Institutionalism—What It Is and What It Hopes to Become; An Approach to World Economics; International Industrial Relations—Migration of Enterprise and Policies Affecting It; World-Wide Depression of 1930; Present Depression—A Tentative Diagnosis; Power and Propaganda; Failure of Electric Light and Power Regulation and Some Proposed Remedies. Pp. 302.

\*\* Supplement No. 2.—Handbook of the Association, 1931.

#### Volume XXII, 1932

The American Economic Review, March, June, September, and December Supplement.—Forty-fourth Annual Meeting:

Papers and Proceedings. Private Enterprise in Economic History; Shorter Working Time and Unemployment; Quantitative Economics; Theory of Technological Progress and the Dislocation of Employment; Measurement of Productivity Changes and the Displacement of Labor; Stabilization of Business and Employment; Principle of Planning and the Institution of Laissez Faire; Institutional Economics; Elasticity of Demand as a Useful Marketing Concept; Investments of Life Insurance Companies; Real Estate in the Business Cycle; Investments and National Policy of the United States in Latin America; Recent Changes in the Character of Bank Liabilities and the Problem of Bank Reserves; Bank Failures in the United States; Transportation by Rail and Otherwise; Our Changing Transportation System. Pp. 306.

#### Volume XXIII, 1933

The American Economic Review, March, June, September, and December Supplement.—Forty-fifth Annual Meeting:

Papers and Proceedings. Rise of Monopoly in the United States; Record of Insurance in the Depression; Some Theoretical Aspects of Unemployment Reserves; The Economics of Unemployment Relief; American Economic Thought; Formation of Capital; Measurement and Relation to Economic Instability; Size of Business Unit as a Factor in Efficiency of Marketing; Reserve Bank Policy and Economic Planning; Federal Reserve Policy in World Monetary Chaos; Tariff Reform: The Case for Bargaining; Speculation in Suburban Lands; Real Estate Speculation and the Depression. Pp. 206.

\*\* Supplement No. 2.—Handbook of the Association, 1933.

#### Volume XXIV, 1934

The American Economic Review, \*\*March, June, September, and December

\*\* Supplement.—Forty-sixth Annual Meeting:

Papers and Proceedings. The History of Recovery; Public Utilities in the Depression; Imperfect Competition; Fundamentals of a National Transportation Policy; Correlation of Rail and Highway Transportation; Marketing under Recovery Legislation; Economics of the Recovery Act; Measurement of Unemployment; Controlled Inflation; Banking Act of 1933—An Appraisal; Some Statistics on the Gold Situation; The Problem of Tax Delinquency; The Problem of Expanding Governmental Activities; The Economics of Public Works. Pp. 224.

#### Volume XXV, 1935

The American Economic Review, March, June, September, and December Supplement.—Forty-seventh Annual Meeting:

Papers and Proceedings. NRA Examined; Rate-making Problems of TVA; New Deal and the Teaching of Economics; Paths of Economic Change; Business Enterprise and the Organization of Production; Changes in the Character, Structure, and Conditions of Production; International Aspects of Problems of Production and Trade; International Movements of Capital; Our Commercial Banking System; Aspects of Co-ordination and Finance; Some Lessons Drawn from European Experience; Nationalism; Security Regulation and Speculation; Monetary Stabilization from an International

Point of View; Monetary Stabilization from a National Point of View; Decentralization of Population and Industry; Co-ordination of State and Local Finance; Relief Aspects of the New Deal; Unified Program for the Unemployed. Pp. 240.

Volume XXVI, 1936

The American Economic Review, March, June, September, and December Supplement.—Forty-eighth Annual Meeting:

Papers and Proceedings. Some Distinguishing Characteristics of the Current Recovery; Price Theories and Market Realities; Notes on Inflexible Prices; Effect of the Depression upon Earnings and Prices of Regulated and Non-regulated Industries; Size of Plants in Its Relation to Price Control and Price Flexibility; Requisites of Free Competition; Monopolistic Competition and Public Policy; Banking Act of 1935; Recent Legislation and the Banking Situation; Economic Aspects of an Integrated Social Security Program; Capital Formation; Trade Agreements Program and American Agriculture; Founding and Early History of the American Economic Association; Developments in Economic Theory; Federal Revenue Act of 1935; Relations between Federal, State, and Local Finances; Equalization of Local Government Resources; Adjustment to Instability; Transportation Problems; Fifty Years' Developments in Ideas of Human Nature and a Motivation; Institutional Economics; Place of Marginal Economics in a Collectivist System; Problem of Prices and Valuation in the Soviet System; Effects of New Deal Legislation on Industrial Relations; Report of the Fiftieth Anniversary Dinner. Pp. 350.

\*\* Supplement No. 2.—Handbook of the Association, 1936.

Volume XXVII, 1937

The American Economic Review, March, June, September, and December Supplement.—Forty-ninth Annual Meeting:

Papers and Proceedings. Economic Interdependence, Present and Future; Quantitative and Qualitative Changes in International Trade During the Depression; Current Tendencies in Commercial Policy; Trade Problem of the Pacific; Analysis of the Nature of American Public Debts; Limits to Possible Debt Burdens, Federal, State, and Local; Debt Retirement and the Budget; United States Debt—Distribution among Holders and Present Status; Federal-State Unemployment Compensation Provisions of the Social Security Act; Unemployment Relief and Insurance; Economic Problems Arising from Social Security Taxes and Reserves; The Situation of Gold Today in Relation to World Currencies; Mechanisms and Objectives for the Control of Exchange; The Adequacy of Existing Currency Mechanisms Under Varying Circumstances; Present Situation of Inadequate Housing; Financing of Housing; Some Economic Implications of Modern Housing; Managed Currency; A Critique of Federal Personnel Policies as Applied to Professional Social Science Positions; New Opportunities for Economists and Statisticians in Federal Employment; Government Employment as a Professional Career in Economics; Indicia of Recovery; Housing and Housing Research; Distribution of Purchasing Power and Business Fluctuations; Forecast of Power Development; The Possibility of a Scientific Electrical Rate System; Co-ordination of Public and Private Power Interests in European Countries; Recent Developments in the Theory of Speculation; Control of Speculation under the Securities Exchange Act; Unorganized Speculation: the Possibility of Control. Pp. 333.

Volume XXVIII, 1938

The American Economic Review, March, June, September, and December Supplement.—Fiftieth Annual Meeting:

Papers and Proceedings. The Significance of Marxian Economics for Present-day Economic Theory; The Significance of Marxian Economics for Current Trends of Governmental Policy; The Rate of Interest; Security Markets and the Investment Process; Relation of Price Policy to Fluctuations of Investment; General Interest Theory; Rate of Interest; Security Regulation; Corporate Price Policies; Fiscal Policies; Rate of Consumption; Wage Rates; Social Security Program; Rate of Consumption; Durable Consumers Goods; Wage Policies. Pp. 192.

Supplement No. 2.—Handbook of the Association, 1938.

## Volume XXIX, 1939

The American Economic Review, \*\*March, \*\*June, September, and December

## Supplement.—Fifty-first Annual Meeting:

Papers and Proceedings. Problem of Industrial Growth in a Mature Economy; Effects of Current and Prospective Technological Developments upon Capital Formation; Public Investment in the United States; Expansion and Contraction in the American Economy; Effect of Industrial and Technological Developments upon Demand for Capital; Role of Public Investment and Consumer Capital Formation; Income and Capital Formation; Price and Production Policies of Large-Scale Enterprise; Changing Distribution Channels; Financial Control of Large-Scale Enterprise; Pure Theory of Production; Changing Character of American Industrial Relations; Wages and Hours in Relation to Innovations and Capital Formation; Effect of Wage Increase upon Employment; Relation of Wage Policies and Price Policies; An Appraisal of Factors Which Stopped Short the Recovery Development in the United States; Fiscal Policy in the Business Cycle; An Appraisal of the Workability of Compensatory Devices; Divergencies in the Development of Recovery in Various Countries; Factors Making for Change in Character of Business Cycle; Industrial Relations. Pp. 280.

## Volume XXX, 1940

The American Economic Review, \*\*March, \*\*June, September, and December

## \*\*Supplement.—Fifty-second Annual Meeting:

Papers and Proceedings. Objectives of Monetary Policy; Economic Issues in Social Security Policy; Bank Deposits and the Business Cycle; Problems in the Teaching of Economics; Price Control Under "Fair Trade" Legislation; Problems of American Commercial Policy; Transportation Problem; Preserving Competition Versus Regulating Monopoly; Theory of International Trade; Collective Bargaining and Job Security; Banking Reform Through Supervisory Standards; Incidence of Taxation; Economic Planning; Growth of Rigidity in Business; Economics of War; Population Problems; Cost Functions and Their Relation to Imperfect Competition. Pp. 436.

## Supplement No. 2—Handbook of the Association, 1940.

No. 5 (February, 1941)

## Fifty-third Annual Meeting (December, 1940):

Papers and Proceedings. Gold and the Monetary System; Economic Research; Federal Budget; Economic Consequences of Deficit Financing; Teaching of Economics; Agricultural Situation; A Review of Fundamental Factors, an Evaluation of Public Measures, and an Appraisal of Prospects; Status and Role of Private Investment in the American Economy, 1940; Unemployment in the United States, 1930-50; Economic Consequences of War Since 1790; Some Economic Problems of War, Defense, and Postwar Reconstruction; United States in the World Economy, 1940; International Economic Relations and Problems of Commercial Policy; Price Policy and Price Behavior. Pp. 458.

## Volume XXXI, 1941

The American Economic Review, March, \*\*June, \*\*September, and December

## Volume XXXII, 1942

The American Economic Review, March, \*\*June, \*\*September, and \*\*December;

## \*\*Supplement.—Fifty-fourth Annual Meeting:

Papers and Proceedings. Economic Adjustments After Wars; Problems of Taxation; Determinants of Investment Decisions; Problems of International Economic Policy for the United States; History of American Corporations; Problems of Labor Market Research; Co-ordination of Federal, State, and Local Fiscal Policy; Technical Aspects of Applying a Dismissal Wage to Defense Workers; Problems of International Economic Policy; Impact of National Defense and the War upon Public Utilities; Future of

Interest Rates; Effect of Managerial Policy upon the Structure of American Business; Economic Effects of Wars; Economic Aspects of Reorganization Under the Chandler Act; Economics of Industrial Research; Objectives in Applied Land Economics Curricula; Changing Position of the Banking System and Its Implications for Monetary Policy; Determination of Wages; Economic Problems of American Cities; Cost and Demand Functions of the Individual Firm; Problems of Price Control; Effects of the War and Defense Program upon Economic Conditions and Institutions; Trade Unions and the Law. Pp. 534.

Supplement No. 2.—Papers Relating to the Temporary National Economic Committee. Pp. 135.

Supplement No. 3.—Directory. Pp. 198.

#### Volume XXXIII, 1943

The American Economic Review, \*\*March, \*\*June, September, and December

\*\*Supplement.—Fifty-fifth Annual Meeting:

Papers and Proceedings. Economic Claims of Government and of Private Enterprise; Our Industrial Plant When Peace Comes; Financial and Government Contract Adjustments of Industry at the End of the War; Problems of Public Policy Raised by Collective Bargaining; Our Labor Force When Peace Comes; Price Control and Rationing; Case Studies in Price Control; Restoration of International Trade; Future of International Investment; International Financial Relations After the War; Economic Regionalism and Multilateral Trade; Bases of International Economic Relations; International Commodity Agreements. Pp. 508 + 15.

#### Volume XXXIV, 1944

The American Economic Review, \*\*March, June, September, and December

\*\*Supplement.—Fifty-sixth Annual Meeting:

Papers and Proceedings. Political Science, Political Economy, and Values; Educational Function of Economists and Political Scientists; Public Administration of Transportation under War Conditions; How Achieve Full and Stable Employment; Incentive Problems in Regulated Capitalism; Postwar Labor Problems; Social Security; Postwar Legal and Economic Position of American Women; Postwar Domestic Monetary Problems; Economic Organization of Welfare; International Trade; Regional Problems; International Monetary Problems. Pp. 440 + 16.

Supplement No. 2.—Implemental Aspects of Public Finance. Pp. 138.

#### Volume XXXV, 1945

The American Economic Review, \*\*March, June, September, and December

Supplement (May).—Fifty-seventh Annual Meeting:

Papers and Proceedings. Consumption Economics; Expanding Civilian Production and Employment After the War; Natural Resources and International Policy; Interdepartmental Courses in the Social Sciences; Price Control and Rationing in the War-Peace Transition; Organized Labor and the Public Interest; Aviation in the Postwar World; International Monetary and Credit Arrangements; Agricultural Price Supports and Their Consequences; Political Economy of International Cartels; Fiscal Problems of Transition and Peace; Problems of Regionalism in the United States; Food and Agriculture—Outlook and Policy; Function of Government in the Postwar American Economy. Pp. 520 + 16.

#### Volume XXXVI, 1946

The American Economic Review, \*\*March, \*\*June, \*\*September, and \*\*December

\*\*Supplement (May).—Fifty-eighth Annual Meeting:

Papers and Proceedings. Problem of "Full Employment"; American Economy in the Interwar Period; Postwar Labor Relations; Monetary Policy; Changing Structure of the American Economy; Economic Problems of Foreign Areas; Publication of an Annual Review of Economics; New Frontiers in Economic Thought; Postwar Shipping Policy; Monopoly and Competition; Postwar Tax Policy; Postwar Railroad Problems; International Investment; Recent Developments in Public Utility Regulation; International Cartels; Economic Research; Methods of Focusing Economic

**Opinion on Questions of Public Policy (e.g., Monetary, Agricultural Price Supports); Undergraduate Teaching of Economics. Pp. 960.**

**\*\*Supplement No. 2—Handbook. Pp. 143.**

**Volume XXXVII, 1947**

**The American Economic Review, \*\*March, \*\*June, September, and December Supplement (May).—Fifty-ninth Annual Meeting:**

Papers and Proceedings. Employment Act of 1946 and a System of National Bookkeeping; Social and Economic Significance of Atomic Energy; Public Debt: History, Effects on Institutions and Income, and Monetary Aspects; Economic Forecasts; Role of Social Security in a Stable Prosperity; Economic Outlook; Economy of the U.S.S.R.; Domestic versus International Economic Equilibrium; Prices: Wartime Heritage and Some Present Problems; Banking Problems; Productivity in the American Economy; International Trade Organization; Vital Problems in Labor Economics; Transportation and Public Utilities Problems; Housing Problems; Economic Research; Changing Character of Money. Pp. 781.

**Volume XXXVIII, 1948**

**The American Economic Review, \*\*March, \*\*June, \*\*September, and December**

**\*\*Supplement (May).—Sixtieth Annual Meeting:**

Papers and Proceedings. Economic Theory of Imperfect Competition, Oligopoly, and Monopoly; Role of Monopoly in the Colonial Trade and Expansion of Europe; Progress of Concentration in Industry; Does Large-Scale Enterprise Result in Lower Costs; Sherman Act and the Enforcement of Competition; Patent Policy; A Consideration of the Economic and Monetary Theories of J. M. Keynes; Keynesian Economics: The Propensity to Consume and the Multiplier, and Savings, Investment, and Wage Rates; Economics Collides with Ethics; An Appraisal of the Taft-Hartley Act; Fiscal Policy in Prosperity and Depression; Problems of Timing and Administering Fiscal Policy in Prosperity and Depression; Transportation and Public Utilities; Futility of Trust-Busting; National Productivity. Pp. 591.

**Volume XXXIX, 1949**

**The American Economic Review, \*\*March, \*\*June, September, and \*\*December**

**Supplement (No. 1—January).—Directory. Pp. 343:**

**Supplement (No. 3—May).—Sixty-first Annual Meeting:**

Papers and Proceedings. Commemoration of the Centenary of the Communist Manifesto—The Sociology and Economics of Class Conflict; Current Research in Business Cycles; Interregional Variations in Economic Fluctuations; Economic Research; Economic Consequences of Some Recent Antitrust Decisions; Theory and Measurement of Price Expectations; Input-Output Analysis and Its Use in Peace and War Economies; Liquidity and Uncertainty; Problems of the ITO; Commodity Marketing—Going Where; John Stuart Mill—Centennial Appraisal; Possibilities for a Realistic Theory of Entrepreneurship; Economics of Preparedness for War; Present Issues of the Latin-American Economy. Pp. 537

**Volume XL, 1950**

**The American Economic Review, \*\*March, \*\*June, September, and December**

**Supplement (No. 2—May).—Sixty-second Annual Meeting:**

Papers and Proceedings. What Planning and How Much Is Compatible with a Market Economy—Recent European Experience; Capitalism and Monopolistic Competition—I. The Theory of Oligopoly, II. Can the American Economy Be Made More Competitive; Capitalism and Economic Progress; Stabilizing the Economy—The Employment Act of 1946 in Operation; Problems of an Advanced Defense Economy; Transportation in Capitalist and Socialized Economies; Can Capitalism Dispense with Free Labor Markets; Capitalism and Equality of Income; Tax Structure and



Private Enterprise; Economic Power Blocs and American Capitalism; American Capitalism—Where Are We Going; U. S. Foreign Investment in Underdeveloped Areas; Economic Policy in Occupied Germany. Pp. 650.  
 Supplement.—Supplement to 1948 Directory. Pp. 41.  
 Supplement No. 2.—The Teaching of Undergraduate Economics. Pp. 226.

#### Volume XLI, 1951

The American Economic Review, \*\*March, \*\*June, September, and December:  
 Supplement (No. 2—May).—Sixty-third Annual Meeting:

Papers and Proceedings. Role and Interests of the Consumer; Institutional Economics; Executive Decisions at the Top Level; Economic Theory, Statistics, and Economic Practice; Capital and Interest; Economic Stabilization; Modern Industrialism and Human Values; Factors in Modern Industrial Development; Government Action or Private Enterprise in River Valley Development; Economic Progress—Some Cases, Comparisons, and Contrasts; Point Four—Development of Backward Areas; Current Problems in International Trade; Economy of the Soviet Union; Changes in National Transportation Policy; Public Utilities and National Policy; Housing Problem—Current Situation and Long-run Effects of Government Housing Programs; Economics of Medical Care—The Problem and Alternative Solutions; Economics in General Education. Pp. 816.

Supplement.—Supplement to 1948 Directory. Pp. 17.

#### Volume XLII, 1952

The American Economic Review, \*\*March, June, September, and December:  
 Supplement (No. 2—May).—Sixty-fourth Annual Meeting:

Papers and Proceedings. Economic Theory and Public Policy; Issues in Methodology; Business Cycle Theory; Monetary Theory; Fiscal Theory; International Trade Theory; Value Theory; Recent Developments in United States Monetary Policy; Inflation Control in the United States; American Foreign Aid Programs; International Trade in the Postwar World; Wages, Manpower, and Rearmament; Governmental Policy on Business Practices; Economic Problems of Military Mobilization; Theoretical Analysis of Economic Growth; General Factors in Economic Growth in the United States; Growth in Underdeveloped Countries; Role of War in American Economic Development; Public Utilities, Transportation, and Spatial Organization; Collective Bargaining in the Regulated Industries. Pp. 768.

Supplement.—Supplement to 1948 Directory. Pp. 11.

#### Volume XLIII, 1953

The American Economic Review, March, June, September, and December:  
 Supplement (No. 2—May).—Sixty-fifth Annual Meeting:

Papers and Proceedings: Monetary Policy; A Stock-taking of Bretton Woods Objectives; Underdeveloped Countries—The Theory and Practice of Technical Assistance; United States Demand for Imports; Interregional Analysis and Regional Development; Governmental Economic Activity; Development of Economic Thought; Technology; Economics in the Curricula of Agricultural Colleges in Canada and the United States; Sociology and Economics; Recent Developments in Mathematical Economics and Econometrics; Research; Long-run Effects of Full Employment on the Labor Market; Theory of Income Distribution; Distribution of Government Burdens and Benefits; Distribution and Utilization of Natural Gas. Pp. 612.

Supplement No. 2.—Graduate Education in Economics. Pp. 223.

Supplement No. 3.—Handbook. Pp. 187.

#### Volume XLIV, 1954

The American Economic Review, March, June, September, and December:  
 \*\*Supplement (No. 2—May).—Sixty-sixth Annual Meeting:

Papers and Proceedings: Fundamental Characteristics of the American Economy; An Appraisal of Economic Change; Factor Markets Versus Product Markets; Farm Prices and Farm Incomes in American Agriculture; Industrial Pricing; Technological Progress and Economic Institutions;

Growth Decisions in the American Economy; Diminishing Inequality in Personal Income Distribution; Wage Determination in the American Economy; Alternative Possibilities of Inflationary Pressures and Higher Cost Bottlenecks in an Economy of Large Bargaining Units and of Less Than Pure and Perfect Competition in the Marketing of Products; Regional Wage Differentials in an Economy of Large Bargaining Units and Less Than Pure and Perfect Competition in the Marketing of Products; Automaticity of Full Employment Under the Assumption of Diminished Defense Expenditures; Institutional Aspects of Savings and Investment; Role of Corporate Taxation in the American Economy; Theory of International Trade in a World of Trade Barriers and Controls and of Variegated National Economic Systems; Economic Doctrines Implied in the Reports of the U.N. and IBRD on Underdeveloped Countries; Corporate International Investment Policies and Programs; Economic Implications of an Aging Population; Report on Graduate Training in Economics; Economics in General Education; Economic and Regulatory Problems in the Broadcasting Industry; National Transportation Policy, Pp. 765.

Supplement.—Supplement to 1953 Handbook. Pp. 11.

#### Volume XLV, 1955

The American Economic Review, \*\*March, June, September, and December

\*\*Supplement (No. 2—May).—Sixty-seventh Annual Meeting:

Papers and Proceedings: International Flow of Economic Ideas; Development Policy in Underdeveloped Countries; Economic Development—Case Studies; Regional Economics; Urbanization and Industrialization of the Labor Force in a Developing Economy; Models of Economic Growth; Current Problems in Agricultural Economics; Long-term Trends in International Trade; Economic Research and Public Policy; Economic Stabilization, Forecasting, and the Political Process; Cyclical Experience in the Postwar Period; Debt Management and Monetary Policy; Taxation and Income Distribution; Concepts of Competition and Monopoly; Impact of Antitrust Laws; Research on the Business Firm; Price and Wage Flexibility; Pricing in Transportation and Public Utilities. Pp. 711.

Supplement.—Supplement to 1953 Handbook. Pp. 8.

#### Volume XLVI, 1956

The American Economic Review, \*\*March, \*\*June, \*\*September, and December

\*\*Supplement (No. 2—May).—Sixty-eighth Annual Meeting:

Papers and Proceedings: Statement of the Problem of Keeping the U.S. Economy Moving Forward, But Steadily; Production and Consumption Economics of Economic Growth; Income Distribution Aspects of Expanding Production and Consumption; Increase of Consumption Part of Economic Growth; Government Expenditures and Economic Growth; Monetary Role in Balanced Economic Growth; Shortening Work Week as a Component of Economic Growth; Highway Development and Financing; Unemployment as a Phase of Economic Growth; Urban Growth and Development; Population Specter—Rapidly Declining Death Rates in Densely Populated Countries; Current Economic Thought and Its Application and Methodology in Continental Europe; Economic Thought and Its Application and Methodology in the East; Economic Potentials of Latin America; Economic Potentials of Africa; Report of the Attorney General's Committee on Antitrust Policy; Changing Patterns of Competition in Transportation and Other Utility Lines; Economics in the Curricula of Schools of Business. Pp. 651.

#### Volume XLVII, 1957

The American Economic Review, March, June, September, and \*\*December

Supplement (No. 2—May).—Sixty-ninth Annual Meeting:

Papers and Proceedings: Institutional Economics; Economic Growth and Development; Keynesian Economics after Twenty Years; Employment Act in the Economic Thinking of Our Times; Business Fluctuations and Inflation; Government Taxing and Spending; Developments in Banking and Monetary Policy; Price and Competitive Aspects of the Distributive Trades;

Monopoly Problem as Seen by Social Scientists; Impact of Some New Developments in Economic Theory; Economics and Changing Technology; Power and Public Utility Problems; Transportation Problems; Social Security and Welfare; Income Distribution; Consumers in the American Economy; International Economics; Soviet Economic Developments; Economics in Schools. Pp. 754.

Supplement (No. 4—July).—Handbook. Pp. 522.†

† This issue is not included in the price of the volume.

#### Volume XLVIII, 1958

The American Economic Review, \*\*March, \*\*June, \*\*September, and December; each \$1.50

\*\*Supplement (No. 2—May).—Seventieth Annual Meeting:

Papers and Proceedings: Veblen Centenary Round Table; Democracy and Trade-Unionism; Agenda for a National Monetary Commission; Is Another Major Business Contraction Likely; Monetary Analysis and the Flow of Funds; Economic Projections and a Statistical Contribution to Price Theory; Statistical Cost Functions; Trends in Capital Investment and Capacity; Income and Consumption; State and Local Public Finance; Agricultural Parity; Further Explorations in Monopolistic-Competitive Price Theory; Petroleum and Natural Gas and the Public Interest; Critical Evaluation of Public Regulation of Independent Commissions; Current Economic Questions Relating to Western Europe; Measuring Production in the U.S.S.R.; Selected Papers—American Economic Association Competition. Pp. 677.

3.00

Supplement.—Supplement to 1956 Handbook. Pp. 32.

1.00

#### Volume XLIX, 1959

The American Economic Review, March, June, \*\*September, and December; each 1.50

Supplement (No. 2—May).—Seventy-first Annual Meeting:

Papers and Proceedings: Soviet Economic Trends and Prospects; Soviet Economic Planning; Non-Russian Communist Economies; Fundamentals of Economic Progress in Underdeveloped Countries; Special Problems Facing Underdeveloped Countries; Role and Character of Foreign Aid; International Trade and Payments in an Era of Coexistence; Maintaining Full Employment and Economic Stability; Balanced Economic Growth in History—A Critique; Economics of Government Expenditures; Power Blocs and the Operation of Economic Forces; Administered Prices Reconsidered; Studies in the Classical Economics; Selected Problems in Economic Theory; Open Competition; Organization and Financing of Economic Research; Market for Economists. Pp. 689.

3.00

#### Volume L, 1960

The American Economic Review, \*\*March, June, \*\*September, and December; each 1.50

Supplement (No. 2—May).—Seventy-second Annual Meeting:

Papers and Proceedings: Standards for the Performance of Our Economic System; Relations Between Economic Theory and Economic Policy; Incentives and Economic Growth: Changing Roles and Public Policies; Problems of Achieving and Maintaining: a High Rate of Economic Growth, Full Employment, and a Stable Price Level; Problem of Raising Incomes in Lagging Sectors of the Economy; Problem of International Harmony: Economic Policies for a Lasting Peace; Problem of Social Priorities; Investing in Education and Research; Facilitating Movements of Labor Out of Agriculture; Reducing Impediments to Foreign Trade; Reforming the Tax System; Improving the Efficiency of the Transportation and Utilities Systems; Research: on Theory of the Firm, on Income, Consumption, and Savings, and on Economic Development. Pp. 745.

3.00

#### Volume LI, 1961

The American Economic Review, March, June, September, and \*\*December; each 1.50

Supplement (No. 2—May).—Seventy-third Annual Meeting:

Papers and Proceedings: Monetary Theory—New and Old Looks; Macroeconomic Theories of Income Distribution; Capital Theory; Managerial Economics—A New Frontier; Frontiers in Uncertainty Theory—The Evidence of Futures Markets; Distribution Costs—Concepts and Measures;

Antitrust Problems; Economic Analysis of Urban Problems; Public Utilities and Transportation; Wheat—A Permanent Need for a Farm Program; Problems of Economic Instability in Other Countries; Balance of Payments of the United States—Problems and Prospects; Economics and National Security; Economic Development in Mainland China; Influence of Moral and Social Responsibility on Economic Behavior; Economic Education—Challenge to Our Profession. Pp. 675.

\$3.00

## Volume LII, 1962

The American Economic Review, \*\*March, June, \*\*September, and \*\*December; each 1.50  
 \*\*Supplement (No. 2—May).—Seventy-fourth Annual Meeting:

Papers and Proceedings: New Developments in the Theory of the Firm; Problems of Economic Development; Lagging U.S. Growth Rate; International Transmission of Business Cycles—Problems and Policies; Soviet Economic Planning; Economics of Research and Development; Studies in Business Behavior; Economic Behavior of Families; Reappraisal of the Doctrine of Consumer Sovereignty; Report of the Commission on Money and Credit; Tax Problems; Systems of Economic Accounts and Analysis for Urban Regions; Role of Transportation in Economic Development; Transportation Problems in the American Economy; Economics of Water Resource Use; Teaching of Economics; Market for Economist. Pp. 615.

3.00

## Volume LIII, 1963

(\$8.00 a Volume)

The American Economic Review, March, June, September, and December; each 2.00  
 Supplement (No. 2—May).—Seventy-fifth Annual Meeting:

Papers and Proceedings: Richard T. Ely Lecture; Public Policies with Respect to Private Business; International Commodity Stabilization; Conditions of International Monetary Equilibrium; Problems of Regional Integration; Problems of Methodology; Topics in Economic Theory; Industrial Capacity; Tax Reform; Financial Institutions and Monetary Policy—A Re-examination of Their Interrelationship; Defense and Disarmament; Pricing and Resource Allocation in Transportation and Public Utilities; Postwar Growth in the United States in the Light of the Long-swing Hypothesis; Economic Trends and Prospects in the U.S.S.R. and Eastern Europe; Japanese Economic Development; Economic Development and the Population Problem; Economic Education. Pp. 753.

4.00

\*\*Supplement No. 2.—Economics in the Schools. Pp. 27.

.50

## Volume LIV, 1964

The American Economic Review, March, June, September, and December; each 2.00  
 Supplement (No. 1—January).—Handbook. Pp. 464.† 8.00

Supplement (No. 3—May).—Seventy-sixth Annual Meeting:

Papers and Proceedings: Richard T. Ely Lecture; Theory of Monopolistic Competition after Thirty Years; Principles of Efficiency; Efficiency in Agriculture; Efficiency in the Labor Markets; Regulated Industries; Efficiency in the Use of Natural Resources; Efficiency in the Government Sector; Problems in Taxation; Financial Factors in Business Cycles; Reappraisals in American Economic History; Comparative Costs and Economic Development; Canadian-American Economic Relations—Efficiency and Independence; Efficiency of the Soviet Economy; Invited Dissertations, I and II; Efficiency in the Teaching of Economics. Pp. 698.

4.00

Supplement No. 2.—Surveys of Foreign Postwar Developments in Economic Thought. Pp. 55.

1.00

† This issue is not included in the price of the volume.

The titles of papers published in the *Papers and Proceedings*, as well as the titles of articles in the *American Economic Review*, are indexed in the *Reader's Guide to 1953* and the *International Index* thereafter.

## NOTICE

Address all inquiries regarding Vols. I-XLVII (1911-57) of the *American Economic Review* to Walter J. Johnson, Inc., 111 Fifth Avenue, New York 3, New York. Inquiries regarding Vol. XLVIII (1958) and subsequent issues of the *Review* and other publications should be addressed to the American Economic Association office.

\* \* \*

The American Economic Association, founded, among other purposes, for the "encouragement of economic research" and "the encouragement of perfect freedom of economic discussion," has over eleven thousand members, including public and professional men and most of the leading students of political economy in America. Membership dues are eight dollars a year. Each member receives all current reports and publications of the Association. In addition to members, there are some four thousand subscribers to the *American Economic Review*, institutional and individual. Applications for membership and other inquiries should be sent to:

Harold F. Williamson

SECRETARY OF THE AMERICAN ECONOMIC ASSOCIATION  
Northwestern University, 629 Noyes Street, Evanston, Illinois